PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Commissioner
US Department of Commerce
United States Patent and Trademark Office, PCT
2011 South Clark Place Room CP2/5C24
Arlington, VA 22202
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 02 May 2001 (02.05.01)	
International application No. PCT/GB00/02809	**Applicant's or agent's file reference** 9.70664/001
International filing date (day/month/year) 20 July 2000 (20.07.00)	**Priority date** (day/month/year) 20 July 1999 (20.07.99)
Applicant NYGREN, Per-Åke et al	

1. The designated Office is hereby notified of its election made:

   [X] in the demand filed with the International Preliminary Examining Authority on:

   20 February 2001 (20.02.01)

   [ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was
   [ ] was not

   made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO **34, chemin des Colombettes** **1211 Geneva 20, Switzerland** Facsimile No.: (41-22) 740.14.35	Authorized officer S. Mafla Telephone No.: (41-22) 338.83.38

PATENT COOPERATION TREATY

From the
INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

To:

GARDNER, Rebecca
FRANK B. DEHN & CO.
179 Queen Victoria Street
London EC4V 4EL
GRANDE BRETAGNE

FILE 70664/001
17 APR 2001
RECEIVED
ANSD

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Rule 71.1)

	Date of mailing (day/month/year) 11.04.2001
Applicant's or agent's file reference 9.70664/001	**IMPORTANT NOTIFICATION**

International application No.	International filing date (day/month/year)	Priority date (day/month/year)
PCT/GB00/02809	20/07/2000	20/07/1999

Applicant
AFFIBODY TECHNOLOGY SWEDEN AB

1. The applicant is hereby notified that this International Preliminary Examining Authority transmits herewith the international preliminary examination report and its annexes, if any, established on the international application.

2. A copy of the report and its annexes, if any, is being transmitted to the International Bureau for communication to all the elected Offices.

3. Where required by any of the elected Offices, the International Bureau will prepare an English translation of the report (but not of any annexes) and will transmit such translation to those Offices.

4. **REMINDER**

The applicant must enter the national phase before each elected Office by performing certain acts (filing translations and paying national fees) within 30 months from the priority date (or later in some Offices) (Article 39(1)) (see also the reminder sent by the International Bureau with Form PCT/IB/301).

Where a translation of the international application must be furnished to an elected Office, that translation must contain a translation of any annexes to the international preliminary examination report. It is the applicant's responsibility to prepare and furnish such translation directly to each elected Office concerned.

For further details on the applicable time limits and requirements of the elected Offices, see Volume II of the PCT Applicant's Guide.

Name and mailing address of the IPEA/	Authorized officer
European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d	Gallego, A

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference 9.70664/001	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
International application No. PCT/GB00/02809	International filing date *(day/month/year)* 20/07/2000	Priority date *(day/month/year)* 20/07/1999
International Patent Classification (IPC) or national classification and IPC C07K1/00		
Applicant AFFIBODY TECHNOLOGY SWEDEN AB		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 7 sheets, including this cover sheet.

   ☐ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of sheets.

3. This report contains indications relating to the following items:

   I ☒ Basis of the report
   II ☐ Priority
   III ☒ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   IV ☐ Lack of unity of invention
   V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   VI ☐ Certain documents cited
   VII ☒ Certain defects in the international application
   VIII ☐ Certain observations on the international application

Date of submission of the demand 20/02/2001	Date of completion of this report 11.04.2001
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Döpfer, K-P Telephone No. +49 89 2399 8547

Form PCT/IPEA/409 (cover sheet) (January 1994)

14

PATENT COOPERATION TREATY

REC'D 19 APR 2001

WIPO PCT

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference 9.70664/001	**FOR FURTHER ACTION** See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)	
International application No. PCT/GB00/02809	International filing date *(day/month/year)* 20/07/2000	Priority date *(day/month/year)* 20/07/1999
International Patent Classification (IPC) or national classification and IPC C07K1/00		
Applicant AFFIBODY TECHNOLOGY SWEDEN AB		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 7 sheets, including this cover sheet.

   ☐ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of sheets.

3. This report contains indications relating to the following items:

   I ☒ Basis of the report
   II ☐ Priority
   III ☒ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   IV ☐ Lack of unity of invention
   V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   VI ☐ Certain documents cited
   VII ☒ Certain defects in the international application
   VIII ☐ Certain observations on the international application

Date of submission of the demand 20/02/2001	Date of completion of this report 11.04.2001
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Döpfer, K-P Telephone No. +49 89 2399 8547

Form PCT/IPEA/409 (cover sheet) (January 1994)

Re Item I

Basis of the report

Re Item III

Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. No International Search Report has been drawn up for the subject-matter of present claim 13. According to Rule 66(1)(e) PCT no International Preliminary Examination is being carried out for the subject-matter of this claim.

Re Item V

Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Reference is made to the following documents:

   D1: WO 99 35293 A (LYNX THERAPEUTICS) 15 July 1999 (1999-07-15)
   D2: WO 98 54312 A (BABRAHAM INSTITUTE) 3 December 1998 (1998-12-03)

2. *Novelty and Inventive Step (Article 33(2)(3) PCT)*

   The present application relates to a combinatorial method for selecting of (a) desired polypeptide(s) comprising (i) a cell free expression step on a solid support (carrying means for biospecific interaction with at least the desired polypeptide) to produce polypeptides, (ii) a separation step in order to obtain the solid support carrying both the desired polypeptide and the nucleic acid encoding it, and optionally (iii) recovery of the nucleic acid and/or said polypeptide. Furthermore, combinatorial libraries are claimed consisting of polypeptides attached to a solid support and associated with expression products of said nucleic acids.
   The expressed polypeptides are fusion proteins.

D1 discloses a purely DNA-based system which is used in the analysis of gene expression. The method comprises the steps of (i) provision of a reference population of nucleic acid sequences attached to solid phase supports in clonal subpopulations, (ii) provision of a population of polynucleotides of expressed genes from a first cell or tissue source and at least one population of polynucleotides of expressed genes from a different cell or tissue source with a light-generating label different from the label comprised by the polynucleotides of any other source, (iii) competitively hybridising the population of said polynucleotides of expressed genes from each source with the reference nucleic acid population to form duplexes, (iv) detecting the optical signal of the labels of the duplexes attached thereto.
Also claimed are mixtures of microparticles bearing the nucleic acid sequences, i.e. combinatorial libraries of these polynucleotides immobilised on solid phase supports.

D2 makes use of ribosome complexes as selection particles for *in vitro* display and evolution of proteins. The selection of proteins is carried out by binding to a ligand, antigen or antibody, and of subsequently recovering the genetic information encoding the protein or peptide from the selected ribosome complex by reverse transcription and PCT (RT-PCR). The RT-PCR step is carried out directly on the intact ribosome complex. The steps of display, selection and recovery can be repeated in consecutive cycles. The method is exemplified using single-chain antibody constructs as antibody-ribosome-mRNA (ARM) complexes.

Neither D1 nor D2 disclose methods or libraries as presently claimed. Thus, the subject-matter of present claims 1-12, 14-17 is considered novel in view of the prior art cited.

Taking D2 as representing the closest prior art, the problem underlying the present application can regarded as to provide alternative methods an libraries for efficiently screening for desired polypeptides or proteins. The solution is the method comprising the steps of present claim 1.
The teaching of D2 seems obviously to be the starting point for the development of the presently claimed method. But neither D2 alone nor in combination with D1 gives a qualified hint to the particular solution with co-immobilised polypeptides,

obtained by cell free expression, and polynucleotides on the same solid support. This library protocol allows rapid screening with high specificity for desired polypeptides. Thus, inventive step can be acknowledged for the subject-matter of present claims 1-12, 14-17.

3. *Industrial applicability (Article 33(4) PCT*

The subject-matter of present claims 1-12, 14-17 appear to comply with the requirements of industrial applicability as stipulated in Article 33(4) PCT.

Re Item VII
Certain defects in the international application

1. Contrary to the requirements of Rule 5.1(a)(ii) PCT, the relevant background art disclosed in the document D2 is not mentioned in the description, nor is this document identified therein.

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	**FOR FURTHER ACTION** see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
9.70664/001		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/GB 00/02809	20/07/2000	20/07/1999
Applicant		
AFFIBODY TECHNOLOGY SWEDEN AB		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 4 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

a. With regard to the **language**, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

[ ] the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

[ ] contained in the international application in written form.
[ ] filed together with the international application in computer readable form.
[ ] furnished subsequently to this Authority in written form.
[ ] furnished subsequently to this Authority in computer readble form.
[ ] the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
[ ] the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. [X] **Certain claims were found unsearchable** (See Box I).

3. [ ] **Unity of invention is lacking** (see Box II).

4. With regard to the **title**,

[X] the text is approved as submitted by the applicant.
[ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

[X] the text is approved as submitted by the applicant.
[ ] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No. 1

[X] as suggested by the applicant.
[ ] because the applicant failed to suggest a figure.
[ ] because this figure better characterizes the invention.

[ ] None of the figures.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB 00/02809

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 13
because they relate to subject matter not required to be searched by this Authority, namely:
see FURTHER INFORMATION sheet PCT/ISA/210

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Claims Nos.: 13

Claim 13, referring to every peptide or nucleic acid identified by the technique of claim 1 ff is indefinite as it might cover every known peptide and nucleic acid. This is found in contrast with the requirements of Art. 6 and Rule 6 PCT, and consequently no search for this claim has taken place.

INTERNATIONAL SEARCH REPORT

International Application No
P[illegible]/GB 00/02809

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/10 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 99 35293 A (LYNX THERAPEUTICS) 15 July 1999 (1999-07-15) claim 1	1-12, 14-17
A	WO 98 54312 A (BABRAHAM INSTITUTE) 3 December 1998 (1998-12-03) the whole document	1-12, 14-17

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
11 January 2001	24 01. 2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Masturzo, P

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
[illegible]/GB 00/02809

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9935293 A	15-07-1999	AU 2113999 A	26-07-1999
		EP 1054999 A	29-11-2000
		NO 20003531 A	05-09-2000
WO 9854312 A	03-12-1998	AU 725957 B	26-10-2000
		AU 7666698 A	30-12-1998
		EP 0985032 A	15-03-2000

I. Basis of the report

1. With regard to the **elements** of the international application *(Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17))*:

   Description, pages:

   1-38 as originally filed

   Claims, No.:

   1-17 as originally filed

   Drawings, sheets:

   1/14-14/14 as originally filed

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

   These elements were available or furnished to this Authority in the following language: , which is:

   ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).

   ☐ the language of publication of the international application (under Rule 48.3(b)).

   ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

   ☐ contained in the international application in written form.

   ☐ filed together with the international application in computer readable form.

   ☐ furnished subsequently to this Authority in written form.

   ☐ furnished subsequently to this Authority in computer readable form.

   ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

   ☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. The amendments have resulted in the cancellation of:

   ☐ the description, pages:

   ☐ the claims, Nos.:

☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

   (Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)

6. Additional observations, if necessary:

III. Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non-obvious), or to be industrially applicable have not been examined in respect of:

   ☐ the entire international application.

   ☒ claims Nos. 13.

because:

☐ the said international application, or the said claims Nos. relate to the following subject matter which does not require an international preliminary examination (*specify*):

☐ the description, claims or drawings (*indicate particular elements below*) or said claims Nos. are so unclear that no meaningful opinion could be formed (*specify*):

☐ the claims, or said claims Nos. are so inadequately supported by the description that no meaningful opinion could be formed.

☒ no international search report has been established for the said claims Nos. 13.

2. A meaningful international preliminary examination cannot be carried out due to the failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions:

   ☐ the written form has not been furnished or does not comply with the standard.

   ☐ the computer readable form has not been furnished or does not comply with the standard.

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

   Novelty (N) Yes: Claims 1-12,14-17

	No:	Claims	
Inventive step (IS)	Yes:	Claims	1-12,14-17
	No:	Claims	
Industrial applicability (IA)	Yes:	Claims	1-12,14-17
	No:	Claims	

2. Citations and explanations
   see separate sheet

VII. Certain defects in the international application

The following defects in the form or contents of the international application have been noted:
see separate sheet

PCT
D INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/10, C12P 19/34, C12Q 1/68, C12N 15/13 — A1

(11) International Publication Number: WO 98/54312

(43) International Publication Date: 3 December 1998 (03.12.98)

(21) International Application Number: PCT/GB98/01564

(22) International Filing Date: 28 May 1998 (28.05.98)

(30) Priority Data:

9710829.4	28 May 1997 (28.05.97)	GB
9724850.4	26 November 1997 (26.11.97)	GB
9804195.7	28 February 1998 (28.02.98)	GB

(71) Applicant *(for all designated States except US)*: BABRAHAM INSTITUTE [GB/GB]; Babraham Hall, Cambridge CB2 4AT (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: TAUSSIG, Michael, John [GB/GB]; Exmoor Cottage, High Street, Hildersham, Cambridge CB1 6BU (GB). HE, Mingyue [CN/GB]; 27 The Close, Babraham, Cambridge CB2 4AQ (GB).

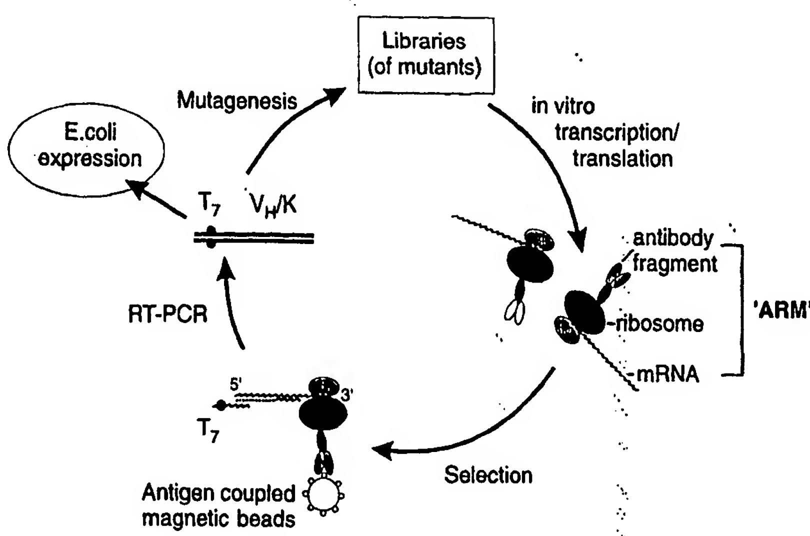
(74) Agent: DAVIES, Jonathan, Mark; Reddie & Grose, 16 Theobalds Road, London WC1X 8PL (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: RIBOSOME COMPLEXES AS SELECTION PARTICLES FOR *IN VITRO* DISPLAY AND EVOLUTION OF PROTEINS



(57) Abstract

The invention provides a method of diplaying nascent proteins or peptides as complexes with eukaryotic ribosomes and the mRNA encoding the protein or peptide following transcription and translation *in vitro*, of further selecting complexes carrying a particular nascent protein or peptide by means of binding to a ligand, antigen or antibody, and of subsequently recovering the genetic information encoding the protein or peptide from the selected ribosome complex by reverse transcription and polymerase chain reaction (RT–PCR). The RT–PCR recovery step is carried out directly on the intact ribosome complex, without prior dissociation to release the mRNA, thus contributing to maximal efficiency and sensitivity. The steps of display, selection and recovery can be repeated in consecutive cycles. The method is exemplified using single–chain antibody constructs as antibody–ribosome–mRNA complexes (ARMs).

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

RIBOSOME COMPLEXES AS SELECTION PARTICLES FOR IN VITRO DISPLAY AND EVOLUTION OF PROTEINS

BACKGROUND TO THE INVENTION

A current focus of interest in molecular biology and biotechnology is in the display of large libraries of proteins and peptides and in means of searching them by affinity selection. The key to genetic exploitation of a selection method is a physical link between individual molecules of the library (phenotype) and the genetic information encoding them (genotype). A number of cell-based methods are available, such as on the surfaces of phages (1), bacteria (2) and animal viruses (3). Of these, the most widely used is phage display. in which proteins or peptides are expressed individually on the surface of phage as fusions to a coat protein, while the same phage particle carries the DNA encoding the protein or peptide. Selection of the phage is achieved through a specific binding reaction involving recognition of the protein or peptide, enabling the particular phage to be isolated and cloned and the DNA for the protein or peptide to be recovered and propagated or expressed.

A particularly desirable application of display technology is the selection of antibody combining sites from combinatorial libraries (4). Screening for high affinity antibodies to specific antigens has been widely carried out by phage display of antibody fragments (4). Combinations of the variable (V) regions of heavy (H) and light (L) chains are displayed on the phage surface and recombinant phage are selected by binding to immobilised antigen. Single-chain (sc) Fv fragments, in which the V_H and V_L domains are linked by a flexible linker peptide, have been widely used to construct such libraries. Another type of single chain antibody fragment is termed V_H/K, in which the V_H domain is linked to the complete light chain, i.e. V_H-linker-V_L-C_L (10). This has a number of advantages, including stability of expression in *E. coli* and the use of the C_L domain as a spacer and as a tag in detection systems such as ELISA and Western blotting. Antibody V_H and V_L region genes are readily obtained by PCR and can be recombined at random to produce large libraries of fragments (21). Such libraries may be obtained from normal or immune B lymphocytes of any mammalian species or constructed artificially from cloned gene fragments with synthetic H-CDR3 regions (third complementarity determining region of the heavy chain) generated *in vitro* (22). Single chain antibody libraries are potentially of a size of

$>10^{10}$ members. Libraries can also be generated by mutagenesis of cloned DNA fragments encoding specific V_H/V_L combinations and screened for mutants having improved properties of affinity or specificity. Mutagenesis is carried out preferably on the CDR regions, and particularly on the highly variable H-CDR3, where the potential number of variants which could be constructed from a region of 10 amino acids is 20^{10} or 10^{13}.

It is clear that for efficient antibody display it is necessary to have a means of producing and selecting from very large libraries. However, the size of the libraries which can potentially be produced exceeds by several orders of magnitude the ability of current technologies to display all the members. Thus, the generation of phage display libraries requires bacterial transformation with DNA, but the low efficiency of DNA uptake by bacteria means that a typical number of transformants which can be obtained is only 10^7-10^9 per transformation. While large phage display repertoires can be created (17), they require many repeated electroporations since transformation cannot be scaled up, making the process tedious or impractical. In addition to the limitations of transformation there are additional factors which reduce library diversity generated with bacteria, e.g. certain antibody fragments may not be secreted, may be proteolysed or form inclusion bodies, leading to the absence of such binding sites from the final library. These considerations apply to all cell-based methods. Thus for libraries with 10^{10} or more members, only a small fraction of the potential library can be displayed and screened using current methodologies. As noted, the size of an antibody library generated either from animal or human B cells or artificially constructed can readily exceed 10^{10} members, while the number of possible peptide sequences encoding a 10 residue sequence is 10^{13}.

In order to avoid these limitations, alternative display systems have been sought, in particular *in vitro* methods which avoid the problem of transformation in library production. One such method is the display of proteins or peptides in nascent form on the surface of ribosomes, such that a stable complex with the encoding mRNA is also formed; the complexes are selected with a ligand for the protein or peptide and the genetic information obtained by reverse transcription of the isolated mRNA.This is known as ribosome or polysome display. A description of such a method is to be found in two US patents, granted to G. Kawasaki/Optein Inc. (16). Therein, semi-random nucleotide sequences (as in a library) are attached to an 'expression unit' and transcribed

in vitro; the resulting mRNAs are translated *in vitro* such that polysomes are produced; polysomes are selected by binding to a substance of interest and then disrupted; the released mRNA is recovered and used to construct cDNA. Two critical parts of the method are the stalling of the ribosome to produce stable complexes, for which cycloheximide is used, and the recovery of the mRNA, for which the bound polysomes are disrupted to release mRNA and the mRNA is then separately recovered. The latter is an integral part of the method as described by Kawasaki and adopted by all others until now. Thus, section VII of the patents (16) deals with the disruption of the polysomes by removal of magnesium, etc; no other method for recovery of RNA or cDNA is suggested other than ribosomal disruption. In US patent no. 5,643,768, claim 1 refers to translating mRNA in such a way as to maintain polysomes with polypeptide chains attached, then contacting to a substance of interest, and finally *isolating mRNA from the polysomes of interest*. In claim 2, cDNA is constructed *subsequent to isolating mRNA from the polysomes that specifically bind to the substance of interest*. This is reiterated in claim 15, wherein step (g) comprises disrupting said polysomes to release said mRNA and step (h) comprises recovering said mRNA, thereby isolating a nucleotide sequence which encodes a polypeptide of interest. Similarly, this is repeated again in claim 29 (e) ... isolating mRNA from the polysomes that specifically react with the substance of interest. In US patent no. 5,658,754, claim 1 (g) also requires disrupting said polysomes to release mRNA; (h) is recovering said mRNA; and (i) is constructing cDNA from said recovered mRNA. However, Kawasaki did not reduce the method to practice in these filings and provided no results. Accordingly, the method was not optimised and he was unaware of the inefficiency of the system as he described it, in particular that due to the method of recovery of mRNA by polysome disruption.

Another description of prokaryotic polysome display, this time reduced to practice, is the international published application WO 95/11922 by Affymax Technologies (18) and the associated publication of Mattheakis et al. (14). Both relate to affinity screening of polysomes displaying nascent peptides, while the patent filing also claims screening of antibody libraries similarly displayed on polysomes. They refer to libraries of polysomes, specifically generated in the *E. coli* S30 system in which transcription and translation are coupled. To produce a population of stalled polysomes, agents such as rifampicin or chloramphenicol, which block prokaryotic translation, are added. The means of recovering the genetic information following

selection of stalled ribosomes is again by elution of the mRNA. In the flowsheet of the method shown in Figure 10 of the patent application (18), an integral part is step 4, namely elution of mRNA from the ribosome complexes prior to cDNA synthesis. The main example in the patent and the publication is of screening a large peptide library with 10^{12} members by polysome display and selection of epitopes by a specific antibody. The polysomes were selected in antibody-coated microplate wells. The bound mRNA was liberated with an elution buffer containing 20mM EDTA and was then phenol extracted and ethanol precipitated in the presence of glycogen and the pellet resuspended in H_2O.

It is clear that the procedures described by Mattheakis et al. are very inefficient at capturing and/or recovering mRNA; thus, on p.72 of the Affymax filing (18), only 1-2% of radiolabelled polysomal mRNA encoding the specific peptide epitope was recovered, which was acknowledged to be low (line 5). The patent application (but not the publication) also includes the selection of an antibody fragment, but with much less detail. In this case, Dynal magnetic beads coated with antigen were used as the affinity matrix. In the example, labelled mRNA was specifically recovered but they did not show recovery of cDNA by RT-PCR. Hence there was no estimation of efficiency or sensitivity, and no demonstration of selection from a library or enrichment.

In a more recent publication (15), Hanes and Pluckthun modified the method of Mattheakis et al. for display and selection of single chain antibody fragments. While retaining the concept, additional features were introduced to make the method more suited to display of whole proteins in the prokaryotic, *E. coli* S30 system. One innovation is the stalling of the ribosome through the absence of a stop codon, which normally signals release of the nascent protein. Once again, recovery of genetic material was by dissociation of the ribosome complexes with 10mM EDTA and isolation of the mRNA by ethanol precipitation (or Rneasy kit) prior to reverse transcription. Separate transcription and translation steps were used, and it was stated that the coupled procedure has lower efficiency; however, no data was provided to this effect. A large input of mRNA was used in each cycle (10μg).

Many additions were incorporated by Hanes and Pluckthun in order to improve the yield of

mRNA after the polysome display cycle, which was initially as low as 0.001% (15). These included stem loop structures at the 5' and 3' ends of the mRNA, vanadyl ribonucleoside complexes as nuclease inhibitor (which also partially inhibit translation), protein disulphide isomerase PDI (which catalyses formation of disulphide bonds) and an anti-sense nucleotide (to inhibit ssrA RNA which in the prokaryotic system otherwises cause the release and degradation of proteins synthesised without a stop codon). The combination of anti-ssrA and PDI improved efficiency by 12-fold overall. However, the yield of mRNA at the end of the cycle, with all additions, was still only 0.2% of input mRNA, expressing the combined efficiency of all steps, including ligand binding (on microtiter wells), RNA release and amplification. Affymax have already described a yield of 2%, i.e. 10-fold higher, as low (cited above).

Hanes and Pluckthun also demonstrated recovery of a specific antibody from a mixture (of two) in which it is initially present at a ratio of $1:10^8$. This required 5 sequential repetitions of the cycle, i.e. using the DNA product of one cycle as the starting point of the next. In Figure 4(A) of ref. 15, there is a considerable carry over of the nonselected polysomes, probably reflecting the method of selection or mRNA recovery. As a consequence, the enrichment factor is relatively low, about 100-fold per cycle.

A further recent ribosome display method was described by Roberts and Szostak (23), in which the nascent protein is caused to bind covalently to its mRNA through a puromycin link. In this system, selection is carried out on these protein-mRNA fusions after dissociation of the ribosome. It thus differs significantly from the other methods described here since it does not involve selection of protein-ribosome-mRNA particles. Its efficiency is only 20-40 fold.

BRIEF DESCRIPTION OF THE INVENTION

It is clear that the described prokaryotic methods of polysome display leave considerable scope for methodological improvement to increase efficiency of recovery of mRNA, sensitivity and selection. In the invention described herein, we have developed a novel, eukaryotic method of ribosome display and demonstrate its application to selection and mutation (evolution) of antibodies and to selection of other proteins from mRNA libraries. It could equally be applied

to isolation of genes from cDNA libraries.

The invention provides a method of displaying nascent proteins or peptides as complexes with eukaryotic ribosomes and the mRNA encoding the protein or peptide following transcription and translation *in vitro*, of further selecting complexes carrying a particular nascent protein or peptide by means of binding to a ligand, antigen or antibody, and of subsequently recovering the genetic information encoding the protein or peptide from the selected ribosome complex by reverse transcription and polymerase chain reaction (RT-PCR). The RT-PCR recovery step is carried out directly on the intact ribosome complex, without prior dissociation to release the mRNA, thus contributing to maximal efficiency and sensitivity. The steps of display, selection and recovery can be repeated in consecutive cycles. The method is exemplified using single-chain antibody constructs as antibody-ribosome-mRNA complexes (ARMs). It is suitable for the construction of very large display libraries, e.g. comprising over 10^{12} complexes, and of efficiently recovering the DNA encoding individual proteins after affinity selection. We provide evidence of highly efficient enrichment, e.g. 10^4- 10^5-fold per cycle, and examples demonstrating its utility in the display and selection of single chain antibody fragments from libraries, antibody engineering, selection of human antibodies and selection of proteins from mRNA libraries.

In its application to antibody fragments, the method is shown in Figure 1. In this form, the method is also termed 'ARM display', since the selection particles consist of antibody-ribosome-mRNA complexes. The antibody is in the form of the single-chain fragment V_H/K described above, but the method is in principle equally applicable to any single chain form, such as scFv. The method differs in a number of particulars from those described above, leading to greater than expected improvements in efficiency, sensitivity and enrichment. In principle, it is based on two experimental results: (i) single-chain antibodies are functionally produced *in vitro* in rabbit reticulocyte lysates (7) and (ii) in the absence of a stop codon, individual nascent proteins remain associated with their corresponding mRNA as stable ternary polypeptide ribosome-mRNA complexes in cell-free systems (8,9). We have applied these findings to a strategy for generating libraries of eukaryotic ARM complexes and have efficiently selected complexes carrying specific combining sites using antigen-coupled magnetic particles. Selection simultaneously captures the relevant genetic information as mRNA.

The coupled transcription/translation system used here is a rabbit reticulocyte extract (Promega) which provides efficient utilisation of DNA. In particular, it avoids the separate isolation of mRNA as described in ref. 15, which is costly in materials and time. The deletion of the stop codon from the encoding DNA is more productive as a means of stalling the ribosome than the use of inhibitors, because it ensures that all mRNA's are read to the 3' end, rather than being stopped at random points in the translation process. The stabilising effect of deletion of the stop codon can be explained by the requirement for release factors which recognise the stop codon and normally terminate translation by causing release of the nascent polypeptide chain (26). In the absence of the stop codon, the nascent chain remains bound to the ribosome and the mRNA. Where it is problematic to engineer stop codon deletion, as in cDNA or mRNA libraries, an alternative method would be the use of suppressor tRNA (charged with an amino acid) which recognises and reads through the stop codon, thereby preventing the action of release factors (24). A further strategy of ribosome stalling would be the use of suppressor tRNA not charged by an amino acid.

In a novel step which introduces a significant difference from preceding methods, we show that cDNA can be generated and amplified by single-step reverse transcription - polymerase chain reaction (RT-PCR) *on the ribosome-bound mRNA*, thus avoiding completely the isolation and subsequent recovery of mRNA by procedures that are costly in terms of material and time. The success and efficiency of this step is surprising, since it is generally assumed that during translation several ribosomes attach to the same mRNA molecule, creating a polysome, and it was not known what effect the presence of several ribosomes in tandem on a single mRNA molecule would have on reverse transcription, where the RT enzyme must read the length of the mRNA. Thus, it is not known whether the enzyme might be able to pass through adjacent ribosomes, or cause their removal from the mRNA, or only function on mRNA molecules to which only one ribosome was attached. Whatever the explanation, this step contributes greatly to the demonstrated efficiency of the system, in which up to 60% of the input mRNA can be recovered in one cycle (Example 6, Figure 9), compared with only 2% in the prokaryotic systems described by Mattheakis et al (14) and 0.2% by Hanes and Pluckthun (15). Furthermore, we have shown that, in the eukaryotic system, extraction of the mRNA from the ribosome complex is five times less effective as a recovery procedure than RT-PCR on the nondisrupted complex and that

much of the mRNA remains bound to the ribosome even after EDTA extraction (Example 8, Figure 11).

The enrichment of individual antibody fragments using ARM display libraries is also more efficient than described for prokaryotic display (15). We have performed experiments which show that mixtures in which the desired specific fragment is present at one part in 10^5 can yield a binding fragment after one cycle, with an effective enrichment factor of $>10^4$ fold, and that cycles can be run sequentially to isolate rarer molecular species from very large libraries (Examples 10 and 11). This is 2-3 orders of magnitude more efficient per cycle than the results reported in the prokaryotic system (15).

Since the ARM libraries are generated wholly by *in vitro* techniques (PCR) and do not require bacterial transformation, their size is limited mainly by the numbers of ribosomes which can be brought into the reaction mixture ($\sim 10^{14}$ per ml in the rabbit reticulocyte kit, according to manufacturer's information) and the amount of DNA which can be handled conveniently per reaction. Hence the production of large libraries becomes much easier than in the phage display method, where the limiting factor is bacterial transformation. An important application is in the selection of proteins from large libraries of mutants; the library can be generated through PCR mutation either randomly or in a site-directed fashion and mutants with required specificity selected by antigen-binding. We demonstrate the use of the ARM display procedure to select antibody (V_H/K) fragments with altered specificity from such libraries. This application to antibody engineering is shown in Example 12, in which the specificity of an anti-progesterone antibody is altered to testosterone binding by a combination of mutagenesis and selection. Such procedures may also be used to produce catalytic antibodies. The operation of the ARM cycle itself also introduces a low level of random mutation through the errors of PCR and we show that the rate of such errors is 0.54% per cycle (Example 9). This can lead to selection of improved properties of affinity and specificity, and is termed 'protein evolution' to indicate the development of novel proteins through a combination of mutation and selection (15). The eukaryotic ARM cycle is well suited to carrying out efficient protein evolution *in vitro*.

The present invention also provides a novel method for obtaining antibodies from libraries made

from immunised mice, bypassing hybridoma technology. In particular, we show that it can be used to make human antibodies by employing a combination of transgenic mouse technology and ARM ribosome display. Mice are available in which transgenic loci encoding human heavy and light chain antibody genes are incoporated into the genomes, such mice giving rise to human antibodies when immunised (20). We provide herein an example in which human antibodies are derived *in vitro* by ARM display of a library prepared from the lymphocytes of such mice (Example 13). This provides a novel route to the derivation of human antibodies for therapeutic purposes.

The ribosome display method described herein is also applicable to any protein or peptide which, having been translated *in vitro*, remains bound to the ribosome and its encoding mRNA. As well as the examples showing the applicability of ARM display to antibodies, we also demonstrate this more general application through translation of an mRNA library obtained directly from normal tissues for selection of individual polypeptide chains (Example 14).

This version of ribosome display thus meets the need for a simple *in vitro* display system for proteins or peptides. It is capable of a very large library size, combined with ease and efficiency of selection and recovery of genetic information; it is also less demanding of special conditions, more sensitive and capable of greater levels of enrichment than methods described hitherto. The combination of a eukaryotic system with efficient mRNA recovery provides a system with a far greater efficiency than would have been predicted by those practiced in the art.

FIGURE LEGENDS

Figure 1. The ARM (antibody-ribosome-mRNA) display cycle, showing the generation of an ARM library by mutagenesis of a single-chain antibody fragment (V_H/K) template, selection of a specific ARM complex by binding to antigen-coupled magnetic beads, and recovery of the genetic information by RT-PCR.

Figure 2A. [SEQ ID 1]. Sequence of the DB3 V_H/K expression construct used in ARM generation. The location of the primers is shown in bold italics. Start points of the V_H, V_L, Cκ

domains and linker are indicated. D1 - D4 are four downstream primers. D1 is used to make the full-length DB3 V_H/K DNA as starting material for the ARM display cycle. D2, D3 and D4 are all recovery primers for use in the first, second and third cycles respectively, in conjunction with the T7 primer (see Figure 3). These primers are suitable for all mouse antibodies with a κ light chain.

Figure 2B. [SEQ ID 2]. Primers used in the modified ARM display cycle. The new upstream T7 primer, including the T7 promoter and protein initiation signal, provide an improved yield. This figure also shows the EVOU primer sequence with the XbaI site underlined. In the recovery phase of the ARM display, the combination of the upstream (T7) primer and both the D2 and EVOU downstream primers lead to recovery of near full length cDNA in each cycle (see Figure 4). These primers are suitable for all mouse antibodies with a κ light chain.

Figure 3. Demonstration that the 3' end of the mRNA is hidden by the ribosome, and that recovery therefore requires the upstream primers D2 and D3 (Figure 2A) for the recovery stages in cycles 1 and 2. In (A), full length DB3 VH/K was transcribed and either primer D1 (1) or D2 (2) used for recovery, which the gel shows was only successful for D2. In (B) the PCR product from cycle A was used in a second cycle with primers D2 (2) or D3 (3); now, the RT-PCR recovery was only successful with primer D3.

Figure 4. Recovery of the same size V_H/K DNA over 5 cycles using the 3-primer method. RT primer = D2 of Figure 2B; PCR primer = EVOU of Figure 2B.

Figure 5. Specific selection of an antibody V_H/K fragment in the ARM cycle.
A. Specific selection of $DB3^R$ ARM complexes by progesterone-BSA-coupled beads. Track 1, RT-PCR of nontranslated $DB3^R$ mRNA selected by progesterone-BSA beads; 2, RT-PCR of $DB3^R$ ARM selected by progesterone-BSA beads; 3, PCR of $DB3^R$ ARM selected by progesterone-BSA beads; 4, RT-PCR of $DB3^R$ ARM selected by testosterone-BSA beads; 5, PCR of $DB3^R$ ARM selected by testosterone-BSA beads; 6, RT-PCR of $DB3^R$ ARM selected by BSA beads; 7, PCR of $DB3^R$ ARMs selected by BSA beads. 8 = 1kb DNA marker.

B. Nonbinding of a $DB3^{H35}$ ARM library to progesterone-BSA-coupled beads. Track 1, 1kb DNA marker; 2, RT-PCR of solution control; 3, RT-PCR of $DB3^{H35}$ ARMs selected by progesterone-BSA beads; 4, RT-PCR of $DB3^{H35}$ ARMs selected by rat anti-κ-coupled beads.

C. Selection of $DB3^{R}$ from ARM libraries containing different ratios of $DB3^{R}$ and $DB3^{H35}$ mutants. Selection was with progesterone-BSA coupled beads. Track 1, ratio of $DB3^{R}$:$DB3^{H35}$ of 1:10; 2, 1:10^2; 3, 1:10^3; 4, 1:10^4; 5, 1:10^5; 6 = $DB3^{H35}$ mutant library alone; 7, 1kb DNA marker

Figure 6. Specific inhibition of the soluble DB3 V_H/K fragment by free steroids in ELISA (right panel), and of DB3 V_H/K in ARM format (centre), demonstrating the same specificity pattern. The centre panel shows the result at 100 ng/ml free steroid. This supports the correct folding of the antibody fragment on the ribosome.

Figure 6A. Effect of DTT (dithiothreitol) concentration in the translation reaction on generation of functional antibody in ARM display.

Messenger RNA encoding DB3 VH/K was generated in an *in vitro* transcription reaction and added to the flexi Rabbit Reticulocyte Lysate system (Promega), which allows DTT to be added separately Track 1. 7: Marker, track 2: untranslated mRNA control, track 3: 0 DTT, track 4: 2mM DTT. track 5: 5 mM DTT, track 6: 10mM DTT. The result shows that 0. 2mM and 5mM DTT all produced good ARM recovery, while only at 10mM was there an inhibition.

Figure 7. Optimisation of Mg^{++} concentration for ARM display.

Figure 8. Optimisation of time course of ARM display.

Figure 9. Efficiency of recovery of input mRNA. cDNA recovered from the ARM cycle (left hand four tracks) is compared with cDNA recovered directly from the mRNA (right hand tracks), in each case by RT-PCR.

Figure 10. Input sensitivity of ARM display, i.e. how little DNA can be used per cycle.

In this experiment, the recovery primer combination was T7 and D4 (Figure 2A). (Note that the original photograph shows a faint but clearly discernable band at 10pg).

Figure 11. Comparison of the method (according to the invention) of recovery of cDNA without ribosome disruption, with that of prior art technology which requires ribosome disruption. The track labelled 'Intact' shows the recovery of cDNA by the present invention, i.e. on the intact ribosome without disruption; 'Disrupted' refers to recovery of cDNA by the prior art method of ribosome disruption using 20mM EDTA and subsequent isolation of mRNA before RT-PCR; and 'Remaining' is recovery of cDNA using the method of the present invention from mRNA remaining associated with the ribosome after disruption according to the prior art method. The relative yields from the 3 recovery reactions was determined by densitometry.

Figure 12. Error rate per cycle. The occurrence of errors during a single cycle of selection of DB3 VH/K ARM was determined by cloning the recovered product after RT-PCR and comparing the sequences of clones with that of the native DB3. Substitutions are highlighted in bold type.

Figure 13. Enrichment of a specific antibody fragment from a library of mutants: analysis by cloning. $DB3^{H35}$ (nonprogesterone-binding) V_H/K was engineered such that the unique HincII site was removed; after ARM selection, treatment with HincII produced a single band of ~800bp. In contrast, similar digestion of $DB3^R$ produces 2 fragments of ~500bp and 300bp. This enables clones containing $DB3^R$ to be distinguished from $DB3^{H35}$ by HincII digestion and gel anaylsis, as shown. $DB3^R$ ARM complexes were selected from mixtures with $DB3^{H35}$ nonbinding mutants at ratios of 1:10 to 1:10^5. The resulting cDNA recovered after one cycle of selection was cloned; DNA was prepared from individual clones and analysed after HincII and EcoRI digestion. In each track, a doublet of bands at 500 and 300bp indicates $DB3^R$ while a single band at ~800bp is $DB3^{H35}$. 10 clones at each ratio were analysed after selection. The result demonstrates an enrichment factor of ~10^4 fold in one cycle. (See Example 10).

Figure 14. Enrichment of $DB3^R$ from a 1:10^6 ratio library ($DB3^R$: $DB3^{H35}$) by repeated ARM display cycles. Selection was with progesterone-BSA coupled beads. Track 1, 1kb DNA marker; 2, RT-PCR after first cycle; 3, RT-PCR after second cycle; 4, RT-PCR after third cycle. The

shortenening of the band between cycles 2 and 3 is due to the use of different primers (D3, D4 respectively).

Figure 15. Changing antibody specificity by mutagenesis and ARM selection (1). DB3 specificity was changed from progesterone-binding to testosterone-binding by mutagenesis of the H-CDR3 loop, followed by a single cycle of ARM selection. Specificity of individual clones was analysed by ARM display, selecting with testosterone-BSA coupled beads. Upper panel: pre-selection clones; lower panel: post-selection clones.

Figure 16. Changing antibody specificity by mutagenesis and ARM selection (2): Selection of DB3 H3 mutants by testosterone-BSA beads in the presence of free progesterone as inhibitor. Track 1: marker; Tracks 2,3: binding of $DB3^R$ to progesterone-BSA (P) or testosterone-BSA (T) beads; Tracks 4,5: binding of the DB3 H3-mutant library to P beads, or to T beads in the presence of free progesterone; Tracks 6,7: the DNA product of track 5 was put into a further ARM display cycle and reselected on P or T beads. (Note the original gel photograph shows a distinct band in track 7).

Figure 17. Changing antibody specificity by mutagenesis and ARM selection (3). Steroid binding of 5 individual clones after selection by testosterone beads was analysed by ARM display and binding to progesterone-BSA beads (P) and testosterone-BSA beads (T).

Figure 18. Changing antibody specificity by mutagenesis and ARM selection (4): Characterisation of a testosterone-specific clone derived by ARM display from the DB3 H3-mutant library. Tracks 1: marker; Tracks 2,3: binding of clone to progesterone-BSA (P) or testosterone-BSA beads (T); Tracks 4,5: binding of clone to T beads in the presence of free progesterone or free testosterone. The sequence of the H3 region of the mutated clone (mut) is shown.

Figure 19. Sequences of human anti-progesterone and anti-testosterone antibodies isolated from an immunised transgenic mouse by ARM display.

Figure 20. Selection of genes from a total mRNA library from mouse spleen cells by ribosome display.

Track 1: Marker

Track 2: RT-PCR of λ light chain on total mRNA from mouse spleen cells.

Track 3: RT-PCR of λ light chain after *in vitro* translation of above mRNA and selection of ribosome complexes by anti-κ coated beads

Track 4: RT-PCR of κ light chain on total mRNA from mouse spleen cells.

Track 5: RT-PCR of κ light chain after *in vitro* translation of the mRNA extract and selection of ribosome complexes by anti-κ coated beads.

Track 6: RT-PCR of Ig heavy chain from total mRNA from mouse B cells.

Track 7: RT-PCR of Ig heavy chain after *in vitro* translation of the mRNA extract and selection of ribosome complexes by anti-κ coated beads.

MATERIALS AND METHOD OF THE ARM RIBOSOME DISPLAY CYCLE (Figure 1)

1. *Single chain antibody constructs used to generate ARM complexes*

The antibody combining sites used to test this method are in a form which we have previously described, namely three-domain single-chain fragments termed V_H/K, in which the heavy chain variable domain (V_H) is linked to the complete light chain (K) (10). We have described a DNA construct and bacterial expression system for producing an anti-progesterone antibody (DB3) as a V_H/K fragment (10) and both periplasmic and cytoplasmic expression were demonstrated (11). The DB3 V_H/K fragment has excellent antigen-binding properties, which in our hands are superior to those of the commonly used single-chain Fv (scFv) form. Using the 'megaprimer' PCR method (12) on plasmid DNA containing DB3 V_H/K, mutants at positions H100 and H35, binding site contact residues for progesterone (13), were produced (unpublished results). $DB3^R$ is a mutant in which tryptophan H100 was substituted by arginine, a modification which leads to an increased affinity for progesterone. $DB3^R$ expressed from *E. coli* bound strongly to progesterone (Ka ~10^9 M^{-1}) but had a much lower affinity for testosterone and none detectable for BSA. In contrast, a library of mutants generated at position H35 (designated $DB3^{H35}$) bound progesterone weakly or not at all. We have employed the $DB3^R$ and $DB3^{H35}$ mutants to test the principle of ARM selection.

2. *Method for generation of ARM complexes*

To generate V_H/K DNA fragments for production of ARMs, PCR was performed using appropriate templates together with (i) an upstream T7 primer, containing the T7 promoter, protein initiation sequence and degenerate sequence complementary to mouse antibody 5' sequences, and (ii) a downstream primer (D1), lacking a stop codon (Figure 2A). The T7 primer sequence was [SEQ ID 3] 5'-gcgcgaatacgactcactatagagggacaaaccatgsaggtcmarctcgagsagtcwgg-3' (where s=c/g, m=a/c, r=a/g and w=a/t), and the D1 primer was [SEQ ID 4] 5'-tgcactggatccaccacactcattcctgttgaagct-3', which contains a BamHI site (underlined) for cloning purposes. To prepare V_H/K constructs, standard PCR was carried out in solution containing 1x PCR reaction buffer (Boehringer Mannheim UK, Lewes, East Sussex), 0.2mM dNTPs (Sigma), 0.3 μM of each primer, 0.05 U/μl of Taq polymerase (Boehringer Mannheim) with one or two drops of nuclease-free mineral oil overlayed on the top of the mixture. The following programme was used: 30 cycles consisting of 94° for 1 min, 54° for 1 min, 72°, for 1 min, then 72° for 10 min followed by 4°.

V_H/K PCR constructs (1ng - 1μg) either purified by QIAquick (QIAGEN) or unpurified, were added to 20μl of the TNT T7 quick coupled transcription/ translation system (Promega UK Ltd, Southampton, Hants SO16 7NS, UK) containing 0.02mM methionine and the mixture incubated at 30° for 60min. The protocol can be scaled down to 10μl. After translation the mixture was diluted with an equal volume of cold phosphate-buffered saline and cooled on ice for 2 min. (For optimisation of conditions, see the description in Examples 4 and 5 below).

3. *Modification of the primers*

The upstream T7 primer, including the T7 promoter and protein initiation signal, can be modified with improved yield. The modified sequence is [SEQ ID 5] 5'-gcagctaatacgactcactataggaacagaccaccatgsaggtcmarctcgagsagtcwgg, as shown in Figure 2B.

4. *Antigen selection of ARM complexes*

Magnetic beads (Dynal, UK) were coupled to bovine serum albumin [BSA], progesterone-11α-BSA, testosterone-3-BSA (Sigma-Aldrich Co. Ltd., Poole, Dorset, UK) or purified rat anti-mouse

κ antibody (gift of Dr G Butcher) according to manufacturer's instructions. 2-3μl of antigen- or anti-κ-conjugated magnetic beads were added to the translation mixture and transferred to 4° for a further 60 min, with gentle vibration to prevent settling. The beads were recovered by magnetic particle concentrator (Dynal MPC), washed 3 times with 50μl cold, sterilised phosphate buffered saline (PBS), pH7.4, containing 0.1% BSA and 5mM magnesium acetate, and once with PBS alone. In order to remove possible DNA contamination, the beads were treated at 37°C for 25 min with DNase I (Promega or Boehringer Mannheim) in 50μl Dnase I buffer (40mM Tris-HCl, pH7.5, 6mM $MgCl_2$, 10mM NaCl, 10mM $CaCl_2$) containing 10 units of enzyme, followed by three washes with 50μl PBS containing 1% tween-20, 5mM magnesium acetate and resuspension in 10μl of diethylpyrocarbonate-treated water.

5. *Recovery and amplification of genetic information from antigen-selected ARM complexes*

To produce and amplify cDNA from the mRNA of antigen-selected ARMs, RT-PCR was performed by adding 2μl of the above bead suspension to 23μl of the RT-PCR mixture (Titan One-tube RT-PCR System, Boehringer Mannheim, or Access RT-PCR system, Promega UK Ltd) containing 1μM of each primer. The primers were the upstream T7 primer described above and a new downstream primer, D2, sequence 5'-cgtgagggtgctgctcatg-3', designed to hybridise at least 60 nt upstream of the 3'-end of ribosome-bound mRNA (Figure 2A). The use of this primer avoids the need to isolate the mRNA from ARM complexes (Figure 1). The reaction mixture was overlayed with one or two drops of nuclease-free mineral oil and placed in a thermal cycler (Techne Progene). The program for single-step RT-PCR was: one cycle at 48° for 45 min, followed by one at 94° for 2 min, then 30-40 cycles consisting of 94° for 30 sec, 54° for 1 min, and 68° for 2 min; finally one cycle at 68° for 7 min was followed by 4°. PCR products were analysed by agarose gel electrophoresis and eluted from the gel for sequencing.

6. *Further cycles of ARM complex generation and selection, and primer combinations for efficient recovery in sequential cycles*

For further cycles, the PCR products produced as above were either gel-purified or added directly to the TNT transcription/translation system. In a second cycle, the RT-PCR downstream primer D3, sequence [SEQ ID 11] 5'-ggggtagaagttgttcaagaag-3', was designed to hybridise upstream of D2 (Figure 2A); similarly in the third cycle the primer D4, [SEQ ID 12] 5'-ctggatggtgggaagatgg-

3', hybridising upstream of D3, was used (Figure 2A). The recovered DNA becomes progressively shorter in each cycle, but a full length V_H/K can be regenerated in any cycle by recombinational PCR. Moreover, the shortening only affects the constant domain of the light chain, not the antigen-binding region.

In this protocol, each cycle required a new downstream primer (D2, D3, D4) due to the fact that the 3' end of the mRNA is covered by the ribosome and inaccessible to primer. While this avoids the need to separate the mRNA from the ribosome, it also causes as noted a shortening of the recovered cDNA in each cycle. We have now overcome this problem by designing a new primer called EVOU, which incorporates D2 and extends downstream. restoring most of the 3' cDNA sequence and which can be used in every cycle.

As is shown in Figure 2B, the sequence of the EVOU primer, is:
5' - **gctcta**gaggcctcacaggtatagctgttatgtcgttcatactcgtccttggtcaacgtg agggtgctgctcat - 3' [SEQ ID 13]
bold = XbaI site

Experiment shows that recovery of cDNA occurs when a mixture of D2 and EVOU are used together in the recovery RT-PCR (Example 1, Figure 4). The unexpected feature of the result is that use of the primer mixture gives just one band of the expected full length whereas two bands were expected. This is probably explained by the efficiency of the EVOU primer under the PCR conditions used, leading to a clean and ideal result.

Therefore, in the preferred method, the primers are the upstream T7 primer and the downstream primer D2, sequence [SEQ ID 14] 5'-cgtgagggtgctgctcatg-3', designed to hybridise at least 60 nt upstream of the 3' end of ribosome-bound mRNA, plus the primer EVOU which incorporates D2, as in Figure 2B.

For further cycles, the PCR products produced as above were either gel-purified or added directly to the TNT transcription/translation system. The combination of D2 and EVOU primers was used in the RT-PCR at the each subsequent cycle. The recovered DNA is thus the same length in each cycle. (Figure 4).

7. Primers for human VH/K antibody fragments

The above primers and those shown in Figure 2 are applicable for VH/K fragments from all mouse immunoglobulins. For human antibodies the corresponding primers are:

T7 primer: 5'-gcagctaatacgactcactataggaacagaccaccatgsaggtmcasctcgagsagtctgg [SEQ ID 6], and

D1 primer: gctctagaacactttcccctgttgaagct [SEQ ID 7]

D2 primer: gctctagagctcagcgtcagggtgctgct [SEQ ID 8]

D4 primer: gctctagagaagacagatggtgcagc [SEQ ID 9]

EVOU primer: cggaattctctagag*tgatggtgatggtgatg*gtagactttgtgtttctcgtagtctgcttt gctcagcgtcagggtgctgct [SEQ ID 10]

(enzyme sites are underlined; hexahistidine tag is in italics).

RESULTS

EXAMPLE 1: RECOVERY OF DNA BY RT-PCR ON THE RIBOSOME COMPLEX AND USE OF 2- OR 3-PRIMER COMBINATIONS

In the ARM method (Figure 1), the ribosome is stalled and the stable complex (nascent protein-ribosome-mRNA) forms because of the absence of a stop codon at the 3' end of the message. Since the ribosome is stalled at the 3' end of the mRNA, the latter should be inaccessible to a 3' primer and/or to reverse transcriptase, necessitating the use of an upstream primer in the recovery of cDNA. This is confirmed by the experiment in Figure 3. When full length DB3 DNA, lacking the 3' stop codon, was transcribed and the mRNA translated *in vitro* and selected with progesterone-BSA beads, cDNA recovery showed that the 3' end of the mRNA was not available for priming in RT-PCR, whereas an upstream primer (D2, Figure 2A) successfully recovered the cDNA. Likewise, in a second cycle, D2 was no longer effective and a primer further upstream (D3, Figure 2A) was required. Thus, the concept of a ribosome bound to the 3' end of the mRNA in the ARM complex appears to be correct. This experiment demonstrates the recovery of cDNA by RT-PCR on the ribosome-mRNA complex.

Clearly, the repeated use of the ARM cycle in this way leads to shortening of the recovered cDNA and eventually it would become necessary to restore full length by a recombinational PCR

reaction. However, in the modified procedure, the use of the D2 primer in combination with the EVOU primer (Figure 2B) restores the full length in every cycle. Figure 4 shows the recovery of the full length VH/K cDNA over 5 cycles. The ARM cycle was performed as described and the combination of primer D2 (labelled as RT primer) and EVOU (PCR primer) was used for recovery. The recovered product DNA was then applied in 4 further sequential cycles in the same way and the products analysed in each case. As shown the full length of VH/K of about 1kb is recovered in each cycle and the DNA was confirmed by sequencing.

The use of these primer combinations leads to efficient recovery of cDNA without the need to isolate the mRNA separately by dissociation of the polysome, as described by others. It is a quick and efficient way of recovering the genetic information as DNA (see also Example 8).

EXAMPLE 2: ANTIGEN-SPECIFIC ARM SELECTION

To demonstrate antigen-specific ARM selection, $DB3^R$ V_H/K was translated *in vitro* and ARMs exposed to magnetic beads coupled either to progesterone-11α-BSA, testosterone-3-BSA or BSA alone. After RT-PCR, a single DNA fragment was detected only from progesterone-11α-BSA coupled beads (Figure 5A, tracks 2,4,6), consistent with the known specificity of $DB3^R$ V_H/K. The recovered fragment was further confirmed as $DB3^R$ by sequencing. No bands were obtained when PCR alone, rather than RT-PCR, was carried out on the progesterone-11α-BSA beads after selection (Figure 5A, tracks 3,5,7), or when the procedure was performed with nontranslated $DB3^R$ mRNA (Figure 5A, track 1). Thus, the band recovered by RT-PCR is derived from mRNA selected via the functional antibody combining site of $DB3^R$ and not from DNA contamination or mRNA carryover.

EXAMPLE 3: INHIBITION BY FREE ANTIGEN OF ARM BINDING TO IMMOBILISED ANTIGEN DEMONSTRATES CORRECT FOLDING OF THE VH/K ON THE RIBOSOME

Inhibition by free steroids can be used to demonstrate the correct folding and functional activity of the ARM complex (Figure 6). The inhibition of DB3 V_H/K expressed as an ARM, using different steroidal inhibitors, is indistinguishable from that of native DB3 and recombinant V_H/K. Furthermore, the 50% inhibition by progesterone-11α-HMS at 1ng (2.5nM) indicates an affinity

very close to that of DB3 (data not shown).

The free steroid inibitors were added to the DB3 ARM mixture in order to block binding to the progesterone-coated beads. They are progesterone-11α-hemisuccinate (HMS) (P11), progesterone-3-carboxymethyloxime (P3); progesterone-6-HMS (P6) and progesterone-21-HMS (P21). The inhibition of free DB3 V_H/K in an ELISA reaction is shown on the right, with the efficiency of the steroids in the order P11>P3>P6>P21. A very similar order of reaction and concentration is seen for the nascent DB3 V_H/K on the ribosome as an ARM (the central panel shows representative results of the recovery RT-PCR reaction).

This demonstration of fine specificity confirms that the nascent antibody V_H/K fragment is correctly folded in the ARM complex. Similarly, there is no requirement for addition of chaperones in the rabbit reticulocyte system, whereas this is also desirable in the prokaryotic system (15). It is possible that the eukaryotic ribosome itself plays a contributory role in folding of the nascent polypeptide chain (25).

EXAMPLE 3A: OPTIMAL DTT CONCENTRATIONS FOR ARM DISPLAY

It has been contended that single chain antibodies may not fold correctly in the presence of 2mM dithiothreitol (DTT), which is present in the transcription/translation reaction mixture, but this appears not to be the case, as shown in Figure 6A. The ARM cycle was carried out in the presence of various concentrations of DTT from 0 - 10mM by translating DB3 VH/K mRNA, produced in a separate transcription in vitro; the translation reaction was performed in the flexi Rabbit Reticulocyte Lysate system (Promega), which allows DTT to be added. The result in Figure 6A shows that 0, 2mM and 5mM DTT all produced good ARM recovery (Tracks 3-5), while only at 10mM was there an inhibition (Track 6). Hence, 2mM DTT does not adversely effect folding and recovery. Thus, protein disulphide isomerase PDI, which is stated as being important for folding of antibody domains in the prokaryotic *E. coli* S30 system (15), is not required for eukaryotic ribosome display in the rabbit reticulocyte system.

EXAMPLE 4: OPTIMISATION OF MAGNESIUM CONCENTRATION (Figure 7)

Magnesium acetate in varying concentrations was added to the TNT transcription/translation reaction system and the recovery of DNA after the ARM cycle was compared. Optimal yield was acheived at 0.5 mM Mg acetate.

EXAMPLE 5: OPTIMISATION OF TIME COURSE (Figure 8)

In the ARM cycle, coupled transcription/translation was carried out for various times in order to determine the optimal time-course of the reaction. This is shown to be 60 minutes incubation, after which time there was no improvement in recovery.

EXAMPLE 6: EFFICIENCY OF RECOVERY OF INPUT MRNA (Figure 9)

In order to assess the efficiency of recovery of mRNA during a single ARM cycle, mRNA for DB3 VH/K was prepared separately by transcription *in vitro*. The cDNA recovered after the processes of translation, ARM complex selection on progesterone beads and RT-PCR on the complexes was compared with that recovered directly from the unmanipulated input mRNA. The left hand 4 tracks show a titration of the cDNA obtained after recovery from the ARM cycle, while the right hand 4 tracks show that obtained from the input mRNA. Densitometry shows that about 60% of the possible cDNA is actually recovered after ARM selection. To produce this result, 60% of the mRNA must be translated into fully functional antigen-binding protein.This recovery yield should be compared with 2% reported by Mattheakis et al. (14) and 0.2% by Hanes and Pluckthun (15) and demonstrates the greatly increased efficiency of the present method.

EXAMPLE 7: SENSITIVITY OF THE ARM CYCLE FOR INPUT DNA (Figure 10)

An essential parameter in the efficiency of the system is the sensitivity for input DNA, i.e. how little DNA can be used per cycle. This experiment, in which DNA input was titrated, shows that a band can be recovered with an input as low as 10pg. The running amount used routinely is 1-10ng (tracks 2 and 3). The sensitivity of the prokaryotic methods by titration is not reported, but the amount used in the Mattheakis method (14) is 440ng and by Hanes and Pluckthun (15) is 10μgm. It is quite likely that the additional steps employed by the latter, namely recovery of

mRNA prior to translation and again prior to reverse transcription, add greatly to the DNA requirement. This can be a critical element in the use of the method to search large libraries. For example, with an input of 1µgm DNA, and a sensitivity of 10pgm, it should be possible to obtain an enrichment of 10^5 fold in a single cycle, which is what we have found (see Example 10). With lower DNA sensitivity, as appears to be the case in the prokaryotic systems, either considerably more DNA would have to be put in, or more selection and recovery cycles carried out.

EXAMPLE 8: COMPARISON OF THE METHOD (ACCORDING TO THE INVENTION) OF RECOVERY OF cDNA WITHOUT RIBOSOME DISRUPTION WITH THAT OF PRIOR ART TECHNOLOGY WHICH REQUIRES RIBOSOME DISRUPTION (Figure 11)

In order to determine the extent to which our procedure for recovery of cDNA at the end of the display cycle, i.e. by RT-PCR on the intact complex, is more efficient than the prior art of Kawasaki (16), Mattheakis (14) and Hanes and Pluckthun (15), we have duplicated their methods by disruption of the ribosome complex and recovery of RNA before RT-PCR. The disruption method followed that described by Hanes and Pluckthun (15): elution buffer was 50mM Tris/ acetate pH7.5, 150mM NaCl, 20mM EDTA; 100µl was added to beads and incubated at 4 °C for 10 min; released RNA was recovered by precipitation with ethanol (standard procedure).

In the gel (Figure 11), the track labelled Intact shows our recovery after one cycle; the track labelled Disrupted is recovery by the disruption method; and track labelled Remaining is what is left behind on the ribosome after disruption. The relative yields were compared by densitometry and showed that recovery performed with the mRNA attached to the ribosome is 5x more efficient than ribosome disruption when applied to the eukaryotic system, and that with the disruption procedure a considerable proportion of the mRNA remains attached to the ribosome and is thus effectively lost. Thus the recovery of cDNA by RT-PCR on the ribosome complex is an important contribution to the increased efficiency of the invention over prior art.

EXAMPLE 9: ACCURACY PER CYCLE (Figure 12)

An important aspect of the invention is its capacity for gradually modifying proteins *in vitro*,

taking advantage of the introduction of random point mutations by the several polymerase reactions included in the cycle followed by ligand-based selection, i.e. protein evolution. At the same time, a very high rate of mutation might render the system nonfunctional by damaging protein structure or combining site specificity. We therefore assessed the errors which are introduced per cycle by cloning the products of an ARM cycle in which DB3 was selected by progesterone-BSA beads. The result in figure 12 shows an error rate of 0.54%, which is low enough to maintain structure but high enough steadily to introduce useful mutations to evolve improved protein capabilities, such as antibody binding site affinity.

EXAMPLE 10: SELECTION OF AN INDIVIDUAL ANTIBODY COMBINING SITE FROM ARM DISPLAY LIBRARIES IN A SINGLE CYCLE. (Figures 5 and 13).

Another important application of ribosome display is the selection of antibodies, or other proteins, from libraries of mutants. To investigate such selection and determine the enrichment possible by eukaryotic ribosome display, $DB3^R$ was mixed with random $DB3^{H35}$ mutants which bind progesterone weakly or not at all (in the mutants, the H35 codon AAC was mutated to C/G T/A/G A). When the $DB3^{H35}$ mutant library alone was displayed as ARM complexes, no DNA band was recoverable after selection with progesterone-11α-BSA beads (Figure 5B, track 3; Figure 5C, track 6); translation of $DB3^{H35}$ was demonstrated by the band obtained with beads coated with rat anti-κ antibody (Figure 5B, track 4). When DNA mixtures containing $DB3^R$ and $DB3^{H35}$ mutants in ratios ranging from 1:10 to $1:10^5$ were displayed as ARMs, a band of V_H/K size was in all cases recovered after a single cycle (Figure 5C, tracks 1-5). Selected DNA was sequenced and , based on codon detection, it was shown that whereas before selection $DB3^R$ was not detectable in the $1:10^3$ - $1:10^5$ libraries, it was the predominant molecule selected from the $1:10^3$ ratio library and a major component of the PCR product from the $1:10^4$ and $1:10^5$ ratio libraries. Thus, enrichment in the range of 10^4 -10^5 fold is achievable in a single cycle of ARM selection.

Because sequencing of a mixed PCR product may not be sufficiently sensitive to provide accurate information on enrichment, in particular to define the ratio of selected : nonselected (background) species, a further means of discriminating between $DB3^R$ and $DB3^{H35}$ mutations

was introduced. A unique HincII enzyme site was removed from $DB3^{H35}$ but left in $DB3^{R}$. Thus, HincII digestion caused a reduction in size of the V_H/K for $DB3^{R}$ from ~800bp to two fragments of ~500bp and 300bp, whereas $DB3^{H35}$ mutants were not cleaved and ran as a fragment of ~800bp. After selection from mixtures in the same ratios as above, the RT-PCR products were cloned and DNA from individual clones mapped by digestion with EcoRI and HincII enabling quantitation of the proportion of $DB3^{R}$ and $DB3^{H35}$ clones recovered. As shown in Figure 13, 70% of the clones selected from a $1{:}10^4$ library and 40% from a $1{:}10^5$ library were $DB3^{R}$. This gives calculated enrichment factors of ~10^4 fold, which is in agreement with the previous data from direct sequencing of PCR mixtures (above). It is possible that even greater enrichment could be obtained by use of larger amount of DNA in the cycle. These enrichment values are considerably higher than those reported for prokaryotic systems of 100-fold (15) or 40-fold (23).

EXAMPLE 11: SELECTION OF AN INDIVIDUAL ANTIBODY COMBINING SITE FROM AN ARM DISPLAY LIBRARY IN TWO OR THREE CYCLES (Figure 14)

While a $1{:}10^6$ $DB3^{R}$:$DB3^{H35}$ library did not produce a detectable RT-PCR band after one cycle (Figure 14, track 2), two further cycles of ARM generation and selection led to recovery of a V_H/K band, with increased intensity at each repetition (Figure 14, tracks 3,4). Sequencing again confirmed the selection of $DB3^{R}$.

EXAMPLE 12: CHANGING ANTIBODY SPECIFICITY BY MUTAGENESIS AND ARM SELECTION FROM A MUTANT LIBRARY (ANTIBODY ENGINEERING) (Figures 15-18)

The affinity of the DB3 antibody for progesterone is ~7,000 times greater than that for testosterone. We attempted to reverse this specificity by combining mutagenesis of the H3 loop (CDR3 of the heavy chain) with ARM display. An H3 mutant library, consisting of 3×10^7 members without stopcodons, was produced by random mutagenesis of $DB3^{R}$ residues 98, 99, 101, 102 and 103. Individual clones from this library, before ARM selection, were analysed by *in vitro* expression in the ARM format as described. In Figure 15, the upper part of the gel (pre-selection clones) shows that there was little or no recovery of cDNA after binding to testosterone-3-BSA-coupled beads. The mutant library was then displayed as ARM complexes and selected in one cycle by binding to testosterone-3-BSA beads. The recovered cDNA was cloned;

individual clones now mostly showed positive binding to testosterone-BSA with strong recovery reflecting good binding (lower part of the gel). This demonstrates that the ARM display method is effective in selective enrichment of mutant clones with new antigen-binding properties and that the ARM system can be used for rapid analysis of binding activity of antibody clones.

The library was then selected against progesterone-BSA and testosterone-BSA beads. For the latter, free progesterone-11α-hemisuccinate was present to block all progesterone binding; hence the effect should be to switch specificity completely to testosterone if such binders are present in the library. In Figure 16, the centre two tracks show this result and demonstrate that the library contains mutants capable of binding specifically to testosterone. The cDNA recovered after binding to testosterone-BSA beads in the presence of free progesterone was recycled against progesterone and testosterone beads and showed specificity for testosterone (tracks 6,7). This result implies that specificity could be switched from binding of one ligand to another. (Note, the band in track 7 is clearly visible on the original photograph).

To confirm the specificity of the cDNA recovered in track 6 of Figure 16, its specificity was also examined by cloning. Figure 17 shows the analysis of individual clones expressed as ARM complexes *in vitro* and tested for binding to progesterone-BSA and testosterone-BSA beads. Out of 5 clones analysed, 3 bound preferentially to testosterone, demonstrating the conversion in specificity from solely progesterone-binding ($DB3^R$) to preferential binding of testosterone (clones 1-3).

One of the clones obtained through mutagenesis and selection against testosterone in the presence of free progesterone was analysed by ARM display and DNA sequencing. In Figure 18, it is seen that the mutant testosterone-binding clone bound specifically to beads coupled to testosterone-3-BSA (T) with no cross-reaction with progesterone-11-BSA (P), and that it could be specifically inhibited by free testosterone-3-BSA (T) but not by free progesterone (P).

These results demonstrate that the ability of ARM display to select from large libraries can be used in conjunction with mutagenesis to carry out antibody engineering, in particular to bring about the alteration of antibody specificity through steps of mutation and selection.

EXAMPLE 13: SELECTION OF HUMAN ANTIBODIES FROM LIBRARIES PREPARED FROM TRANSGENIC MICE. (Figure 19)

An area of great interest is the use of display methods to isolate human antibodies which can be used for *in vivo* diagnostic or therapy in man. The source of such a library can be human lymphocytes from naturally immune or actively immunised individuals. However, in order to respond to human antigens, many of which are important therapeutic targets, the human lymphocytes must develop in a nontolerising environment. This can be achieved through the use of transgenic mice, which have acquired the genes encoding human heavy and light chains in their genomes through embryo manipulation; the ability of these mice to make endogenous mouse antibody has been eliminated by introduction of knock-out deletions (20). Such mice respond to immunisation with human antigens by production of human antibodies (20). We have immunised mice carrying a human heavy chain translocus comprising 5 V_H genes, the complete D-J region and the Cμ and Cδ genes, together with a light chain translocus carrying 8 V_L genes, the entire J region and the Cκ gene. The mice were immunised with progesterone-11α-HMS-BSA and after 8 weeks the spleens were removed. A V_H/K DNA library was prepared by RT-PCR amplification of the expressed V_H and light chain genes followed by random combination through the standard V_H/K linker sequence, using recombinational PCR; the stop codon was deleted from the 3' end of the light chain. The library was expressed *in vitro* as ARM complexes and selected using progesterone-BSA or testosterone-BSA coupled magnetic beads. Recovered cDNA was cloned and sequenced (Figure 19). The sequences enabled human VH and VL genes to be identified and the CDR3 regions of the heavy chain to be compared. While there is repetitive selection of two human VH/VL combinations (VH4/Vk1-12 and VH1-2/Vk4-01) there is considerable diversity in the H3 sequences. However, one of the steroid contact residues identified from crystallography in the VH CDR2 of anti-steroid antibodies (W50, the first CDR2 residue)) is universally present and a relevant aromatic is also often present around residue 100.

EXAMPLE 14. SELECTION OF GENES FROM AN MRNA LIBRARY BY EUKARYOTIC RIBOSOME DISPLAY. Figure 20

Although the examples cited thus far have all related to expression and selection of antibody fragments, ribosome display should be applicable to any protein which retains a selectable

functionality, such as a binding site or an epitope, when bound in nascent form on the ribosome. Thus, it should be possible to isolate genes from cDNA or mRNA libraries in the ribosome display format, e.g. selecting complexes with antibody- or ligand-coupled particles.

This example demonstrates the use of ribosome display (1) to select a gene encoding an expressed protein starting with an mRNA extract obtained from mammalian cells, (2) to select a specific mRNA as a ribosome complex using an antibody attached to beads as the selecting agent, and (3) to recover the relevant gene by RT-PCR carried out on the ribosome-bound mRNA. For the library, mRNA was extracted by Pharmacia mRNA purification kit and directly expressed *in vitro* using the Promega TNT transcription/translation system. No attempt was made to remove the stop codon, but instead the reaction was stopped after 1 hour by cooling on ice. The translation mixture was exposed to monoclonal rat anti-κ antibody linked to magnetic beads. Bound mRNA was converted to cDNA and amplified by RT-PCR using specific primers for the κ chain and, as negative controls, for λ light chain and IgG heavy chain. The results are shown in Figure 20. The cDNA bands in tracks 2, 4 and 6 were obtained directly from the mRNA library and show that mRNA for human λ and κ light chains and heavy chain respectively were present. After the expression of the mRNA in ribosome display format and selection with anti-κ coated beads, a strong κ light chain band was recovered after RT-PCR (track 4), with no band for λ light chain (track 3) and a weak band for heavy chain (track 7), thus demonstrating the specific selection and recovery of κ chain cDNA. To our knowledge, this is the first experiment to show the selection of a protein from a natural library (i.e. derived from a normal tissue) by ribosome display.

CONCLUSIONS

The greater efficiency of this display method over those previously described can be seen as deriving from a number of factors, the use of a eukaryotic expression system, coupled transcription and translation, stalling the ribosome by eliminating the stop codon and efficient recovery by RT-PCR carried out on the ribosome complex. Thus no time or material is consumed in isolating mRNA at different stages (after transcription, after selection) as in the Hanes and Pluckthun description. The novel step is the one of recovery, which we have demonstrated to be superior to ribosome dissociation. It is also likely to be much more economical due to the fact

it allows much smaller amounts of mRNA to be handled in the system, which is clearly important when selecting rare molecular species from large libraries. We have shown that very small amounts of input DNA can be recovered, making it practicable to use large libraries.

REFERENCES

(1) Smith, G.P (1985) *Science* **228**, 1315-1317.

(2) Georgiou, G., Poetschke, H.L., Stathopoulos, C and Francisco, J.A (1993) *TIBTECH* **11**, 6-10.

(3) Kasahara, N., Dozy, A.M. and Kan, Y.W. (1994) *Science* **266**, 1373-1376.

(4) Winter, G., Griffiths, A.D., Hawkins, R.E. and Hoogenboom, H.R. (1994) *Annu. Rev. Immunol.* **12**, 433-455.

(5) Wrighton, N.C., Farrell, F.X., Chang, R., Kashyap, A.K., Barbone, F.P., Mulcahy, L.S., Johnson, D.L., Barrett, R.W., Jolliffe, L.K. and Dower, W.J. (1996) *Science*, **273**, 458-463.

(6) Kretzschmar, T., Zimmermann, C. and Geiser, M. (1995) *Anal. Biochem.* **224**, 413-419.

(7) Nicholls, P.J.,Johnson, V.G., Andrew, S.M., Hoogenboom, H.R., Raus, J.C.M and Youle, R.J.(1993) *J. Biol. Chem.* **268**, 5302-5308.

(8) High, S., Gorlich, D., Wiedmann, M., Rapoport, T.A. and Dobberstein, B. (1991) *J.Cell. Biol.* **113**, 35-44.

(9) Fedorov, A.N & Baldwin, T.O.(1995) *Proc. Natl. Acad. Sci.* **92**, 1227-1231.

(10) He, M. Kang, A.S., Hamon, M., Humphreys, A.S., Gani, M and Taussig, M.J. (1995) *Immunology* **84**, 662-668.

(11) He, M. Hamon, M.Liu, H., Kang, A.S. and Taussig, M.J. (1995) *Nucleic Acids Res.* **23**, 4009-4010.

(12) Sarkar, G and Sommer, S.S (1990) *Biotechniques* **8**, 404-407.

(13) Arevalo, J., Stura, E.A., Taussig, M.J. and Wilson, I.A. (1993) *J. Mol. Biol.* **231**, 103-118.

(14) Mattheakis, L.C., Bhatt, R.R. and Dower, W.J. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 9022-9026.

(15) Hanes, J. and Pluckthun, A. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 4937-4942.

(16) Kawasaki G, US patents 5,643,768 Cell free synthesis and isolation of novel genes and polypeptides (Jul. 1st 1997) and 5,658,754 (Aug 19 1997)

(17) Vaughan et al. (1996) *Nature Biotech.* **14**, 309-314.

(18) PCT publication WO 95/11922, 'In vitro peptide and antibody display libraries', dated 25 October 1994.

(19) Stansfield I., Jones K.M. and Tuite M.F. (1995) *Trends in Biochem. Sci.* **20**, 489-491.

(20) Bruggemann M. and Neuberger M.S. (1996) *Immunology Today* 17, 391-397.

(21) Marks J.D. et al. (1992) *Bio/Technology* **10**, 779-783

(22) Hawkins R.E. et al. (1992) *J. Mol. Biol.* **226**, 889-896.

(23) Roberts R.W. and Szostak J.W. (1997) *Proc. Nat. Acad. Sci USA* **94**, 12297-12302

(24) Laski F.A. et al. (1984) *EMBO Journal* 3, 2445-2452.

(25) Netzer W.J. and Hartl. F.U. (1998) *Trends in Biochem. Sci.* 23, 68-73.

CLAIMS

1. A method for the display and selection of proteins or peptides and for recovery of the genetic material encoding them, which method consists of

(a) transcription and translation of DNA in a cell free system such that complexed particles are formed, each comprising at least one individual nascent protein or peptide or other DNA expression product associated with one or more ribosomes and the specific mRNA encoding the protein or peptide;

(b) contacting the said complexed particles with a ligand, antigen, antibody or other agent in order to select particles through binding to the protein or peptide product, and

(c) recovering the genetic information encoding the protein or peptide as DNA by means of reverse transcription and polymerase chain reaction (RT-PCR) carried out on the mRNA while the latter remains bound to the said complexed particle.

2. A method according to claim 1 in which the transcription/translation systems are eukaryotic.

3. A method according to claims 1 and 2 in which transcription and translation are coupled.

4. A method according to claim 1 in which the transcription/translation system is a rabbit reticulocyte lysate system

5. A method according to claims 1 and 2 which involves making protein (or peptide)-ribosome-mRNA complexes from DNA and mRNA lacking a stop codon.

6. A method according to claim 1(b) wherein the agent selecting the complexed particles is immobilised and bound to magnetic beads, plastic dishes or other insoluble support.

7. A method in which DNA is produced by reverse transcription followed by polymerase chain reaction (RT-PCR), carried out on mRNA physically linked with one or more ribosomes after translation of the mRNA.

8. A method for the display and selection of proteins or peptides and for recovery of the genetic material encoding them, which method consists of
(a) coupled transcription and translation of DNA lacking a stop codon in a cell free rabbit reticulocyte system such that complexed particles are formed comprising at least one individual nascent protein or peptide or other DNA expression product associated with one or more ribosomes and the specific mRNA encoding the protein or peptide;
(b) contacting the said complexed particles with an insolubilised ligand, antigen, antibody or other agent in order to select particles through binding to the protein or peptide product, and
(c) recovering the genetic information encoding the protein or peptide as DNA by means of reverse transcription and polymerase chain reaction (RT-PCR) carried out on the mRNA while the latter remains bound to the said complexed particle.

9. A method according to claims 1, 5 and 8 in which the protein is a single chain antibody fragment.

10. A method according to claim 9 in which the single chain antibody fragment comprises the variable region of the heavy chain (V_H) linked to the variable region of the light chain (V_L) (scFv fragment) or the entire light chain (K) (V_H/K fragment).

11. Primers for carrying out the RT-PCR reaction of the method of claims 1 and 8, to recover antibody fragments from antibody-ribosome-mRNA complexes, such primers being selected from the primers referred to in SEQ ID Nos. 3-14.

12. A method which involves subsequent incorporation of the RT-PCR product DNA obtained by the method of claims 1 and 8 into an expression vector and production of the protein or peptide by transformation of bacteria such as *E. coli*.

13. A display library comprising proteins, peptides or other DNA expression products complexed with eukaryotic ribosomes and the specific mRNAs encoding those proteins, peptides or other products.

14. A display library according to claim 13 in which the mRNA molecules lack stop codons.

15. A protein-ribosome-mRNA display library according to claims 13 or 14 in which the individual members comprise proteins capable of binding specifically to ligands, allowing the subsequent selection of individual members of the library by binding to immobilised ligand.

16. A library according to claims 13 or 14 in which the proteins displayed are antibodies or antibody fragments, including single chain fragments comprising different numbers of domains, such as V_H, V_L, scFV, V_H/K, Fab.

17. A library according to claims 13 or 14 in which the products displayed are receptors.

18. A library according to claims 13 or 14 in which the products displayed are peptides.

19. A library according to claims 13 or 14 in which the products displayed are protein mutants.

20. A library according to claim 16 in which the antibodies or fragments are obtained from lymphocytes of immunised or non-immunised animals or humans.

21. A library according to claims 13 or 14 generated by means of mutation of cloned DNA encoding antibodies, receptors or fragments thereof.

22. A method according to any preceding method claim which involves selection of individual mutants from the display library according to claim 19 or 21.

23. The use of a ribosome display library according claim 18 encoding peptides for identification and mapping of epitopes recognised by specific antibodies or receptors.

24. A method for making antibodies of a mouse, rat or other mammal which consists of
(a) contacting the animal with antigen,
(b) making a DNA library comprising combinations of the V_H and V_L regions of the immunoglobulins of said animal, linked as single chain Fv or V_H/K fragments as in claim 10,

(c) creating a eukaryotic ribosome display library by *in vitro* transcription and translation of said DNA library, such that complexed particles are formed each comprising at least one individual nascent antibody fragment associated with one or more ribosomes and the specific mRNA encoding the antibody fragment,

(d) selecting complexed particles carrying specific antibody fragments through binding to an antigen or other agent,

(e) recovering the genetic information encoding the antibody fragment by means of RT-PCR carried out on the mRNA while bound to the said particle,

(f) expressing and collecting said antibody fragments.

25. A method for making human antibodies which consists of

(a) contacting with antigen a transgenic mouse carrying human loci encoding heavy and/or light chains of immunoglobulins as transgenes,

(b) making a DNA library comprising combinations of the V_H and V_L regions of the human immunoglobulins of said animal, linked as single chain Fv or V_H/K fragments as in claim 10,

(c) creating a eukaryotic ribosome display library by *in vitro* transcription and translation of said DNA library, such that complexed particles are formed each comprising at least one individual nascent antibody fragment associated with one or more ribosomes and the specific mRNA encoding the antibody fragment

(d) selecting such complexed particles carrying specific antibody fragments through binding to an antigen or other agent,

(e) recovering the genetic information encoding the antibody fragment as DNA by means of RT-PCR carried out on the mRNA while bound to the said particle,

(f) expressing and collecting said antibody fragments.

26. A method for the display of proteins or peptides as complexed particles and for recovery of the genetic information encoding them, consisting of

(a) translating mRNA or an mRNA library in a eukaryotic cell free system such that complexed particles are formed, each comprising at least one individual nascent protein or peptide or other expression product associated with one or more ribosomes and the specific mRNA encoding the protein or peptide;

(b) contacting the particles with a ligand, antibody or other agent in order to obtain selection of particles by means of binding to the protein or peptide product, and
(c) recovering the genetic information encoding the product as DNA by means of RT-PCR carried out on the mRNA while bound to the particle.

27. A method for displaying proteins or peptides as complexed particles and for recovery of the genetic information encoding them, consisting of
(a) transcribing and translating cDNA or a cDNA library in a eukaryotic cell free system such that the complexed particles are formed, each comprising at least one individual nascent protein or peptide or other expression product associated with one or more ribosomes and the specific mRNA encoding the protein or peptide;
(b) contacting the said particles with a ligand, antibody or other agent in order to obtain selection of particles by means of binding to the protein or peptide product, and
(c) recovering the genetic information encoding the product by means of reverse transcription and polymerase chain reaction carried out on the mRNA while bound to the particle.

28. The use of repeated cycles of ribosome display and selection according to any preceding method claim.

29. The use of a eukaryotic ribosome display library according to any preceding library claim in a method to select ligands for combining sites or receptors, such ligands having potential uses as drugs or therapeutics.

30. The use of a ribosome display library according to any preceding library claim in a method to isolate genes through binding of translated products to immobilised antibody or ligand.

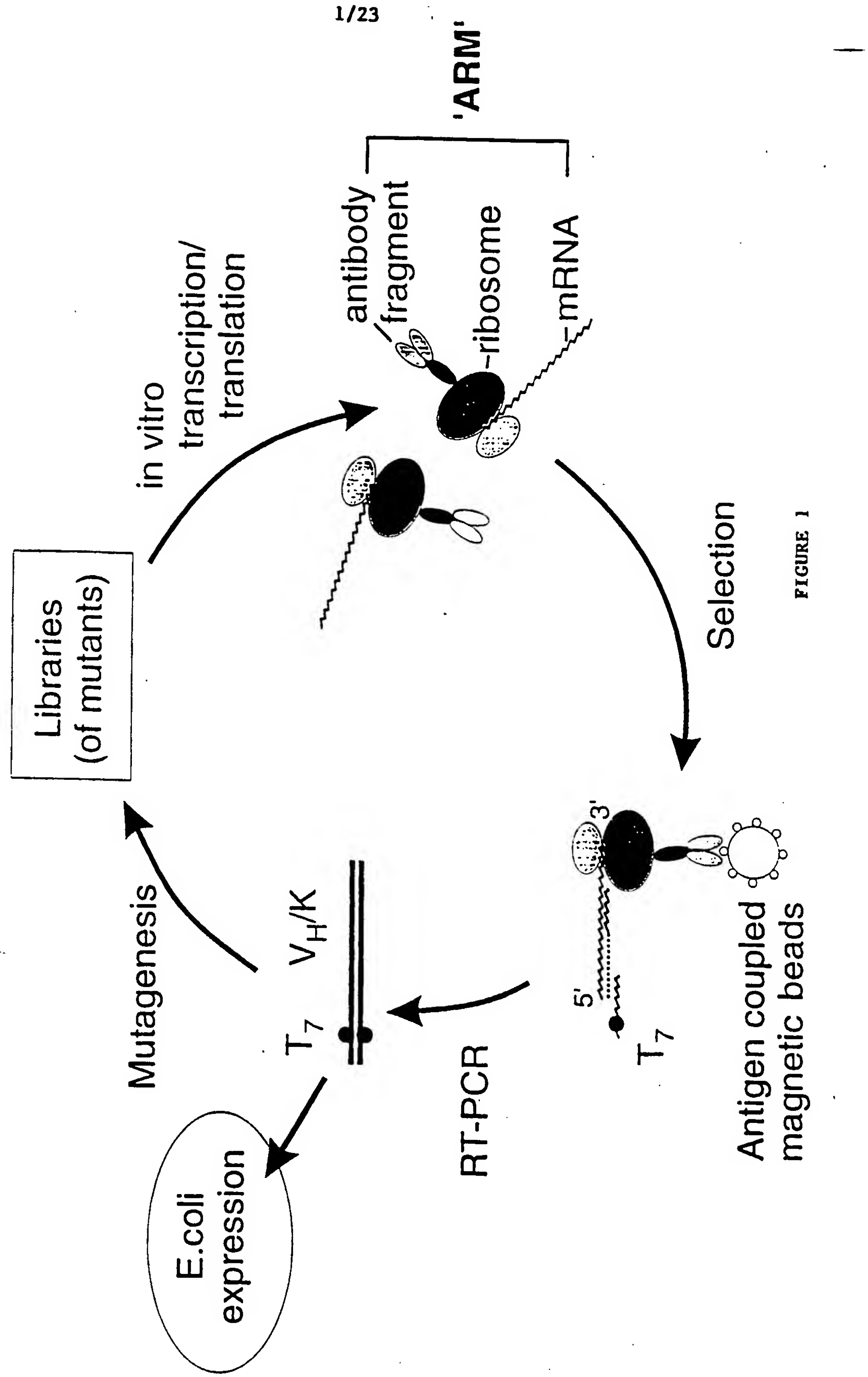

'ARM'
antibody fragment
ribosome
mRNA
in vitro transcription/ translation
Libraries (of mutants)
Selection
3'
5'
T7
Antigen coupled magnetic beads
RT-PCR
V_H/K
T7
Mutagenesis
E.coli expression

FIGURE 1

T7 primer
5'-***gcgcgaatacgactcactatagagggacaaaccatgsaggtcmarctcgagsagtcwgg***
V_H ->

acctgagctgaagaagcctggagagacagtcaagatctcctgcaaggcttctgggtatgcct

tcaaaaactatggagtgaactgggtgaaggaggctccaggaaaggatttaaagtggatgggc
H^{35}

tggataaacatctacactggggagccaacatatgttgatgacttcaagggacggtttgcctt

ctctttggaaacctctgccagcactgcctatttggagatcaacaacctcaaaaatgaagaca

cggcaacgtatttctgtacaagaggtgactacgtcaactggtacttcgatgtctggggcgca
H^{---}

gggaccacggtcaccgtctcctcagccaaaacgacacccccatctgtctatccactggccga
linker ->

gctcgtgatgacccagattccactctccctgcctgtcaatcttggagatcaagcctccatct
V_L ->

cttgcagatctagtcagagccttgtacacagtaatggaaacacctatttacattggtacctg

cagaagccaggccagtctccaaagctcctgatctacaaagtttccaaccgattttatggggt

cccagacaggttcagtggcagtggatcagggacagatttcacactcaagatcagcagagtgg

aggctgaggatctgggaatttatttctgctctcaaagttcacatgttcctcccgacgttcggt

ggaggcaccaagctggaattcaaacgggctgatgctgcaccaactgtatccatcttcccacc
Cκ -> ***ggtagaagggtgg***
D4 primer

atccagtgagcagttaacatctggaggtgcctcagtcgtgtgcttcttgaacaacttctacc
taggtc-5' ***gaagaacttgttgaagatgg***
D3 primer

ccaaagacatcaatgtcaagtggaaaattgatggcagtgaacgacaaaatggcgtcctgaac
gg-5'

agttggactgatcaggacagcaaagacagcacctacagcatgagcagcaccctcacgttgac
gtactcgtcgtgggagtgc-5'
D2 primer

caaggacgagtatgaacgacataacagctatacctgtgaggccactcacaagacatcaactt

cacccattgtcaagagcttcaacaggaatgagtgtggtggatccagtgca-3'
tcgaagttgtccttactcacaccacctaggtcacgt-5'
D1 primer

Figure 2A [SEQ ID 1]

T7 primer

5'-***gcagctaatacgactcactataggaacagaccaccatgsaggtcmarctcgagsagtcwgg***

V_H ->

acctgagctgaagaagcctggagagacagtcaagatctcctgcaaggcttctgggtatgcct

tcaaaaactatggagtgaactgggtgaaggaggctccaggaaaggatttaaagtggatgggc

H^{35}

tggataaacatctacactggggagccaacatatgttgatgacttcaagggacggtttgcctt

ctctttggaaacctctgccagcactgcctatttggagatcaacaacctcaaaaatgaagaca

cggcaacgtatttctgtacaagaggtgactacgtcaaccgttacttcgatgtctggggcgca

H^{100}

gggaccacggtcaccgtctcctcagccaaaacgacacccccatctgtctatccactggcccga

linker ->

gctcgtgatgacccagattccactctccctgcctgtcaatcttggagatcaagcctccatct

V_L ->

cttgcagatctagtcagagccttgtacacagtaatggaaacacctatttacattggtacctg

cagaagccaggccagtctccaaagctcctgatctacaaagtttccaaccgattttatggggt

cccagacaggttcagtggcagtggatcagggacagatttcacactcaagatcagcagagtgg

aggctgaggatctgggaatttatttctgctctcaaagttcacatgttcctccgacgttcggt

ggaggcaccaagctggaattcaaacgggctgatgctgcaccaactgtatccatcttcccacc

Cκ ->

atccagtgagcagttaacatctggaggtgcctcagtcgtgtgcttcttgaacaacttctacc

ccaaagacatcaatgtcaagtggaaaattgatggcagtgaacgacaaaatggcgtcctgaac

agttggactgatcaggacagcaaagacagcacctacagcatgagcagcaccctcacgttgac

D2 primer: gtactcgtcgtgggagtgc-5'

EVOU primer: gtactcgtcgtgggagtgcaactg

caaggacgagtatgaacgacataacagctatacctgtgaggccactcacaagacatcaactt

gttcctgctcatacttgctgtattgtcgatatggacactccggagatctcg-5'

XbaI

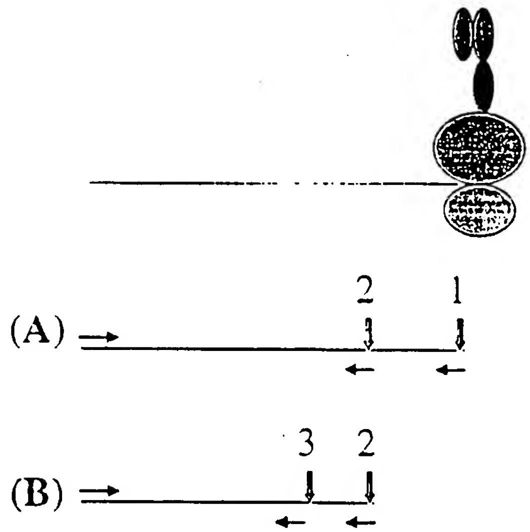
cacccattgtcaagagcttcaacaggaatgagtgtggtggatccagtgca-3'

tcgaagttgtccttactcacaccacctaggtcacgt -5'

D1 primer

Figure 2B [SEQ ID 2]

3' end of ARM mRNA is inaccessible in RT-PCR



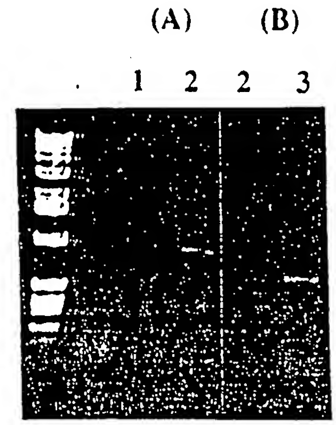


Figure 3

ARM Cycles: Full length DNA recovery by three primer RT-PCR

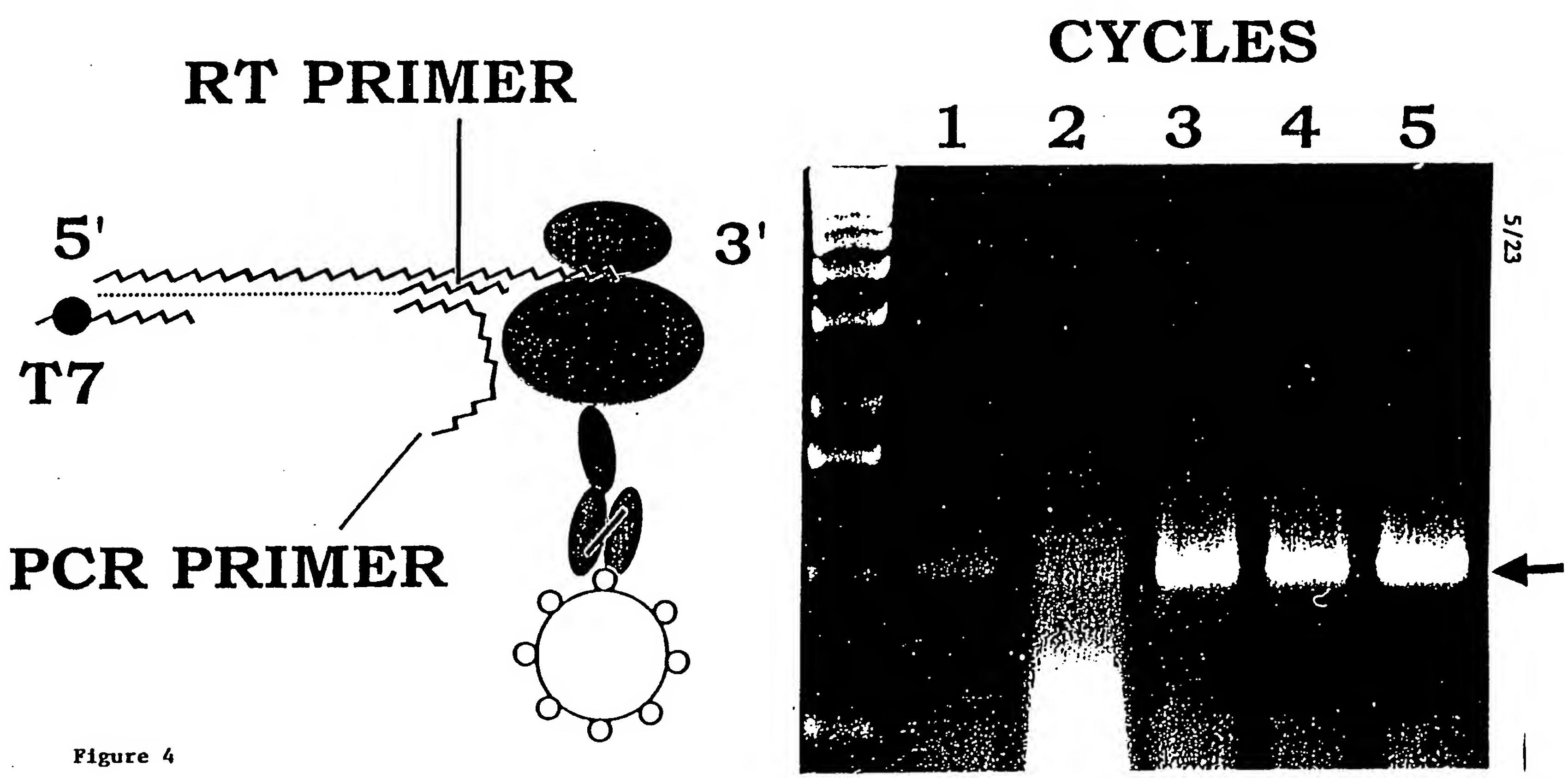


Figure 4

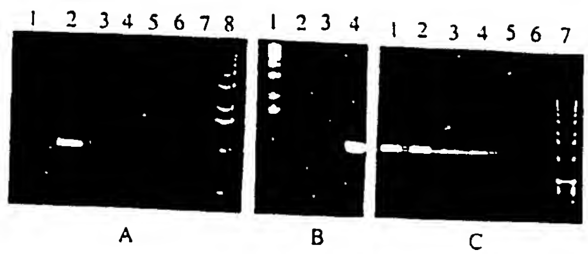


Figure 5

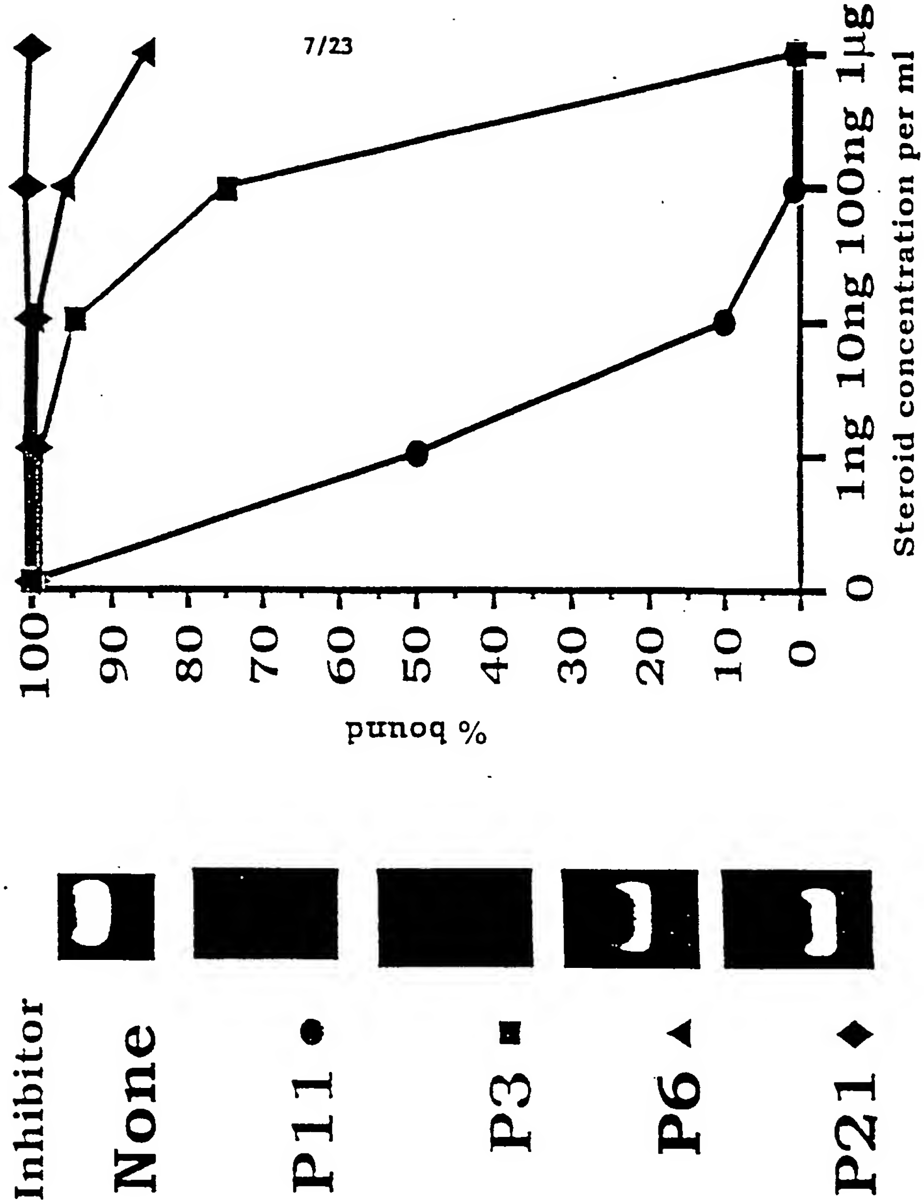

100
90
80
70
60
50
40
30
20
10
0
% bound
0
1ng 10ng 100ng 1μg
Steroid concentration per ml
Inhibitor
None
P11 ●
P3 ■
P6 ▲
P21 ◆

Figure 6

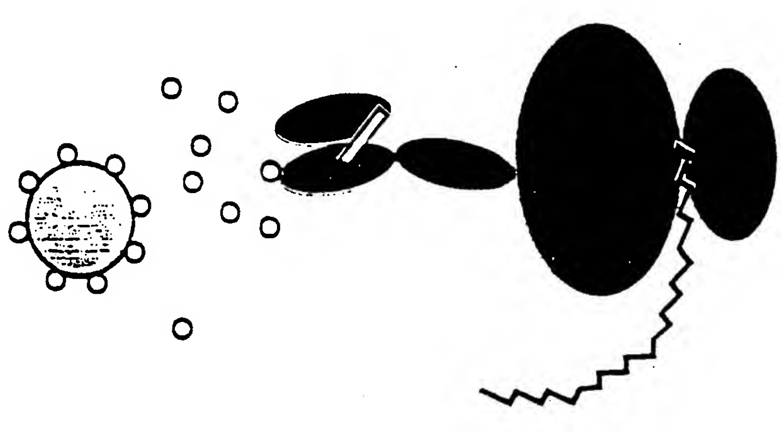


Figure 6A

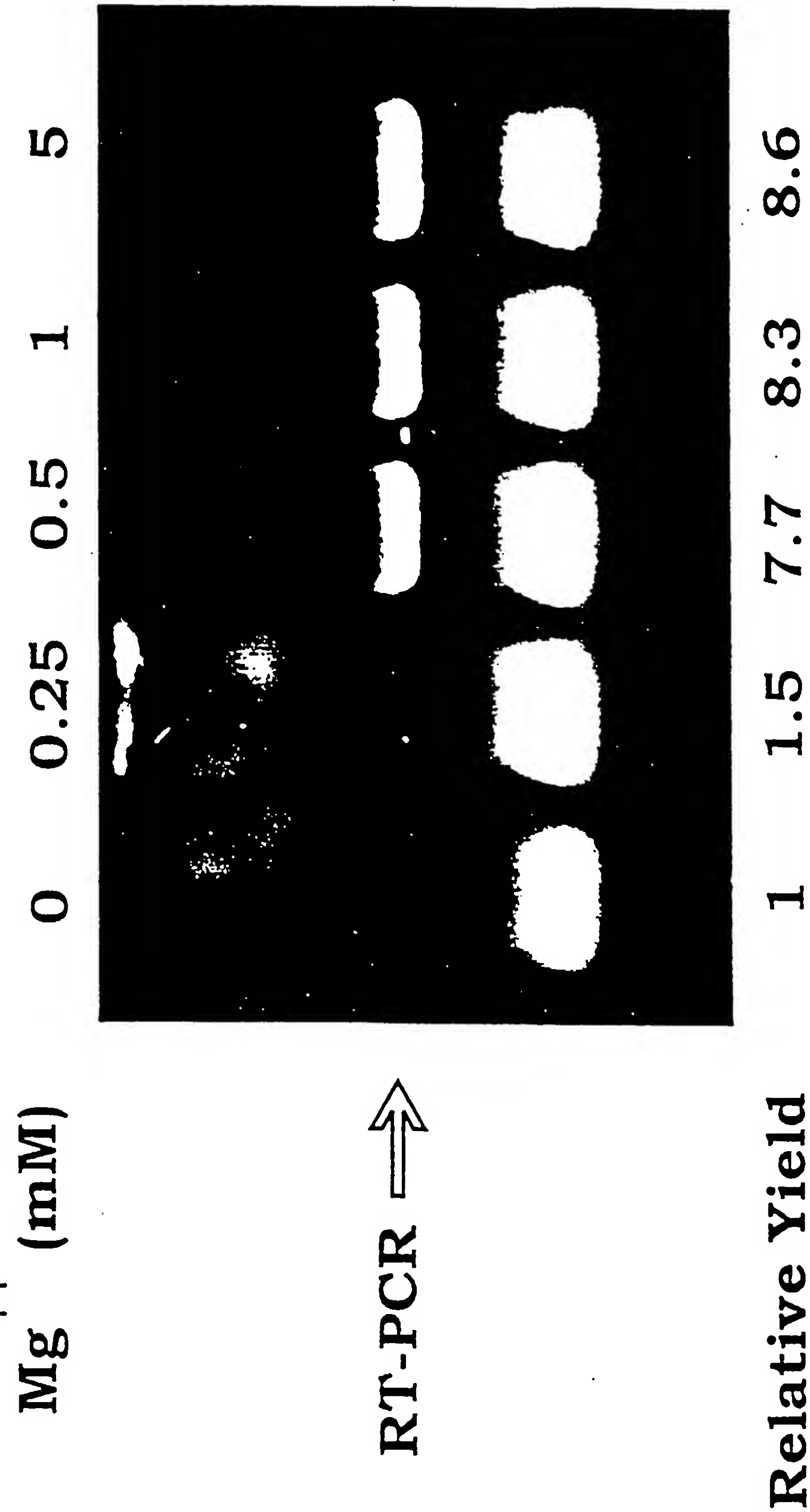


Figure 7

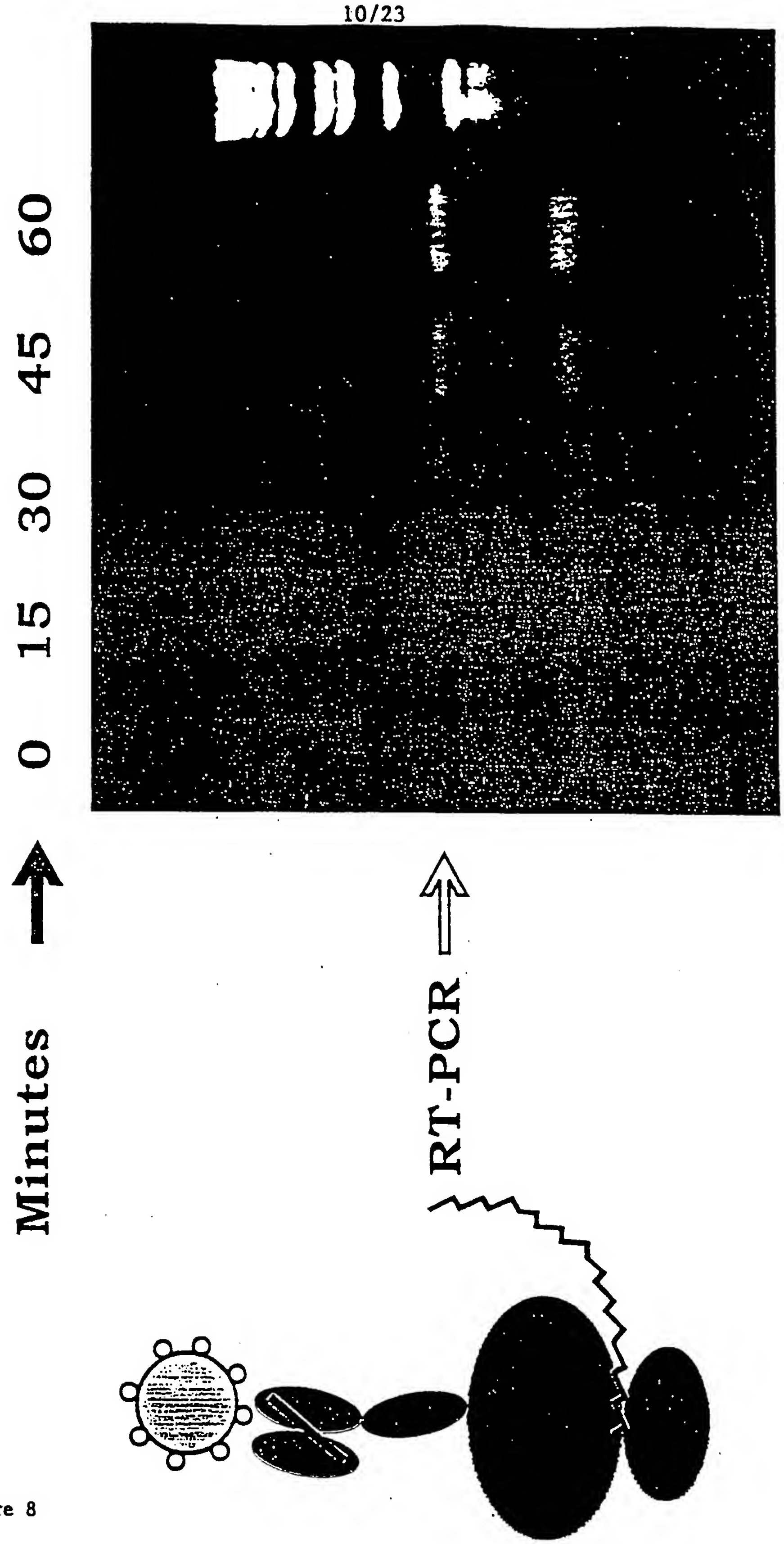


Figure 8

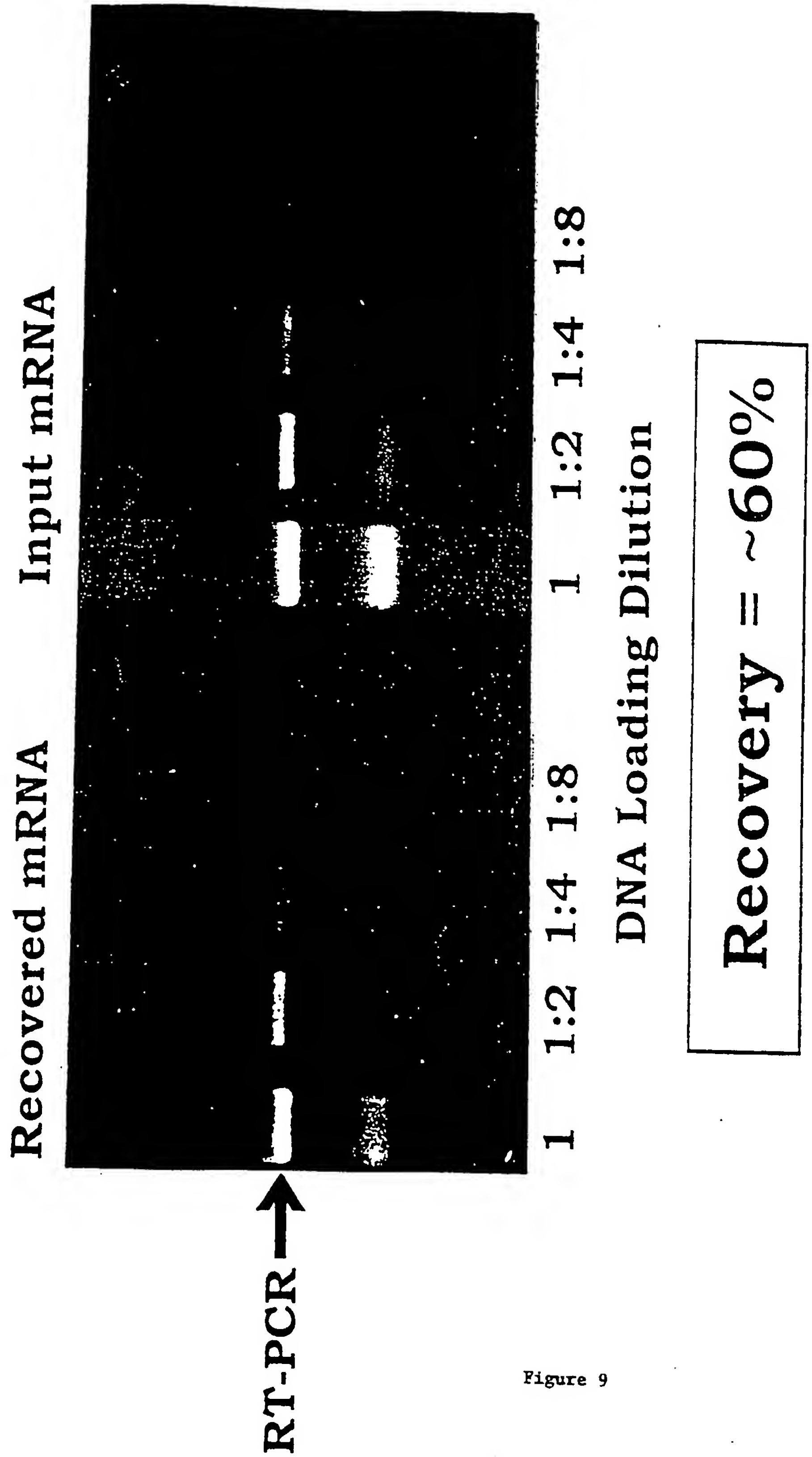

Recovered mRNA
Input mRNA
1 1:2 1:4 1:8 1 1:2 1:4 1:8
DNA Loading Dilution
Recovery = ~60%
RT-PCR

Figure 9

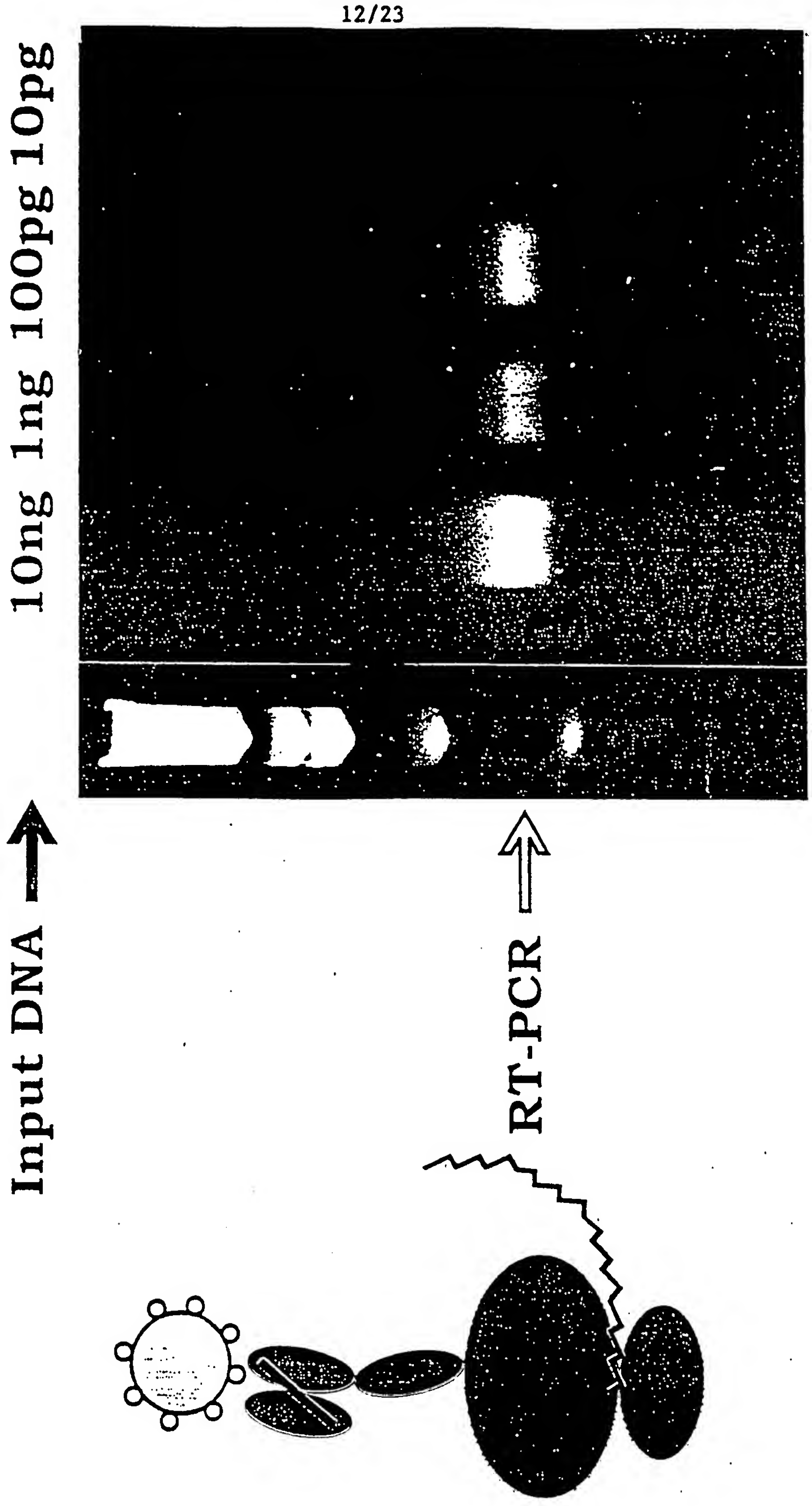


Figure 10

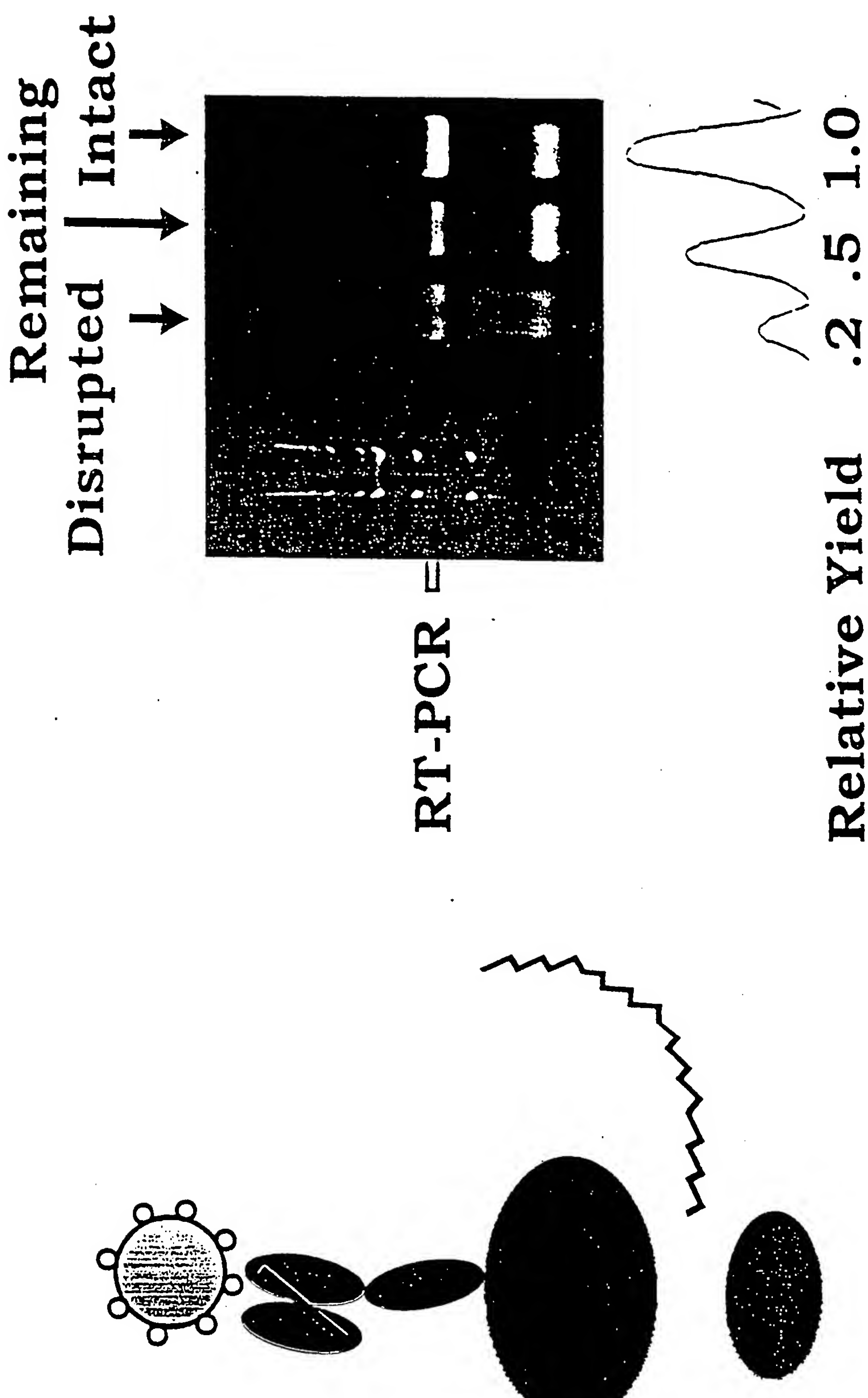


Figure 11

ARM SELECTION: ERROR RATE IN ONE CYCLE

DB3 : QSGPELKKPGETVKISCKASGYAFKNYGVNWVKEAPGKDLKWMGWINIYT

(1) QSGPELKKPGETVTISCKASG**F**AFKNYG**A**NWVKEAPGKDLKWMGWIYIY**S**
(2) QSGPELKKPGETVKISCKASGYAFKNYGVNWVKEAPGKDLKWMGWINIY**S**
(3) QSGPELKKPGETVKISCKASGYAFKNYGVNWVKEAPGKDLKWMGWINIYT
(4) QSGPELKKPGETVKISCKASGYAFKNYG**A**NWVKEAPGKDLKWMGWINIYT
(5) QSGPELKKPGETVKISCKASGYAFKNYGVNWVK**G**APGKDLKWMGWINIYT
(6) QSGPELKKPGETVKISCKASGYAFKNYGVNWVKEAPGKDLKWMGWINIYT

DB3 : GEPTYVDDFKGRFAFSLETSASTAYLEINNLKNEDTATYFCTRGD
(1) GEPTFVDDFKGRFAFSLETSAS...
(2) GEPTYVDDFKGRFAFSLETSASTAYLEI**TY**LKNEDTATYFCTRGD
(3) GEPTYVDDFKGRFAFSLETSASTAYLEINNLKNEDTATYFCTRSD
(4) GEPTYVDDFKGRFAFSLETSASTAYLEINNLKNEDTATYFCTRSD
(5) GEPTYVDDFKGRFAFSLETSASTAYLEINNLKNEDTATYFCTRGD
(6) GEPTYVDDFKGRFAFSLETSASTAYLEI**??**LKNEDTATFFCTRGD

Nucleotide error: 9/1682 =0.54%

Figure 12

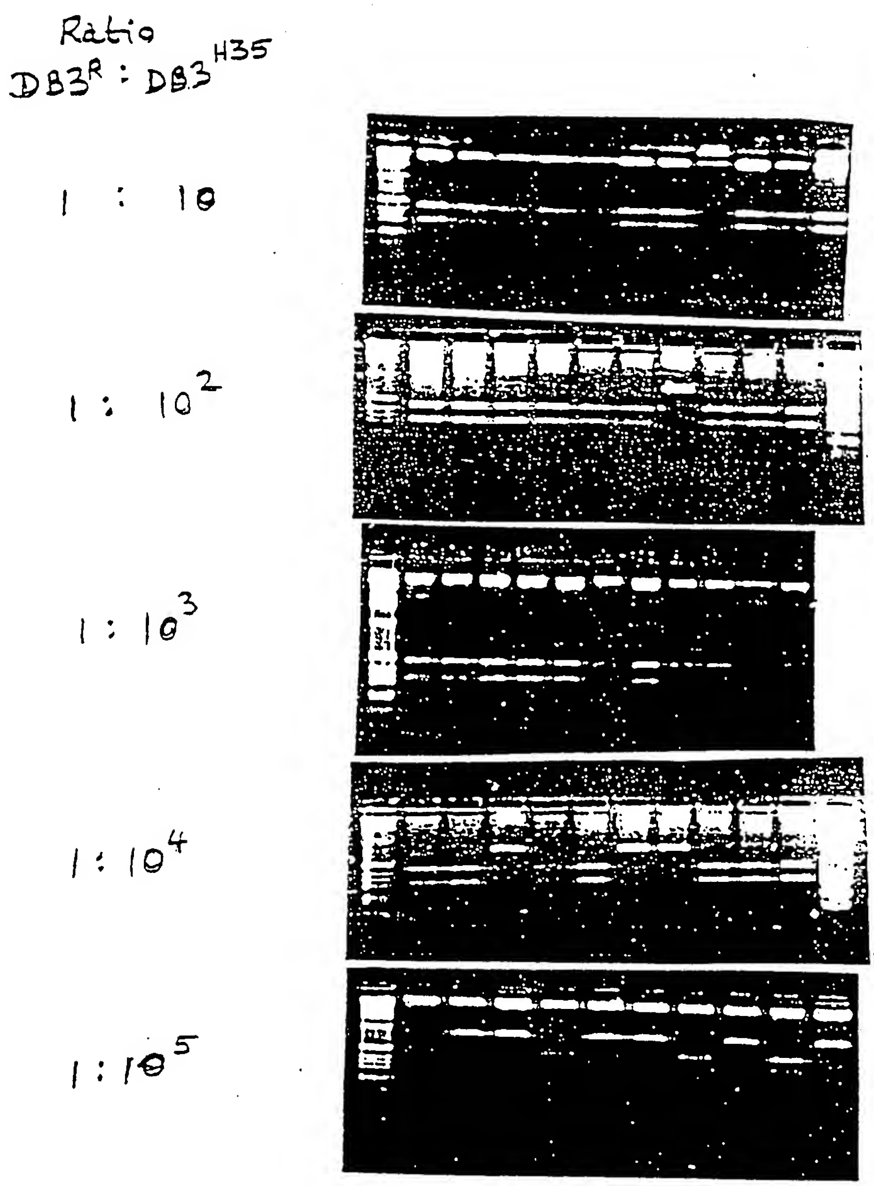


Figure 13

16/23

Figure 14

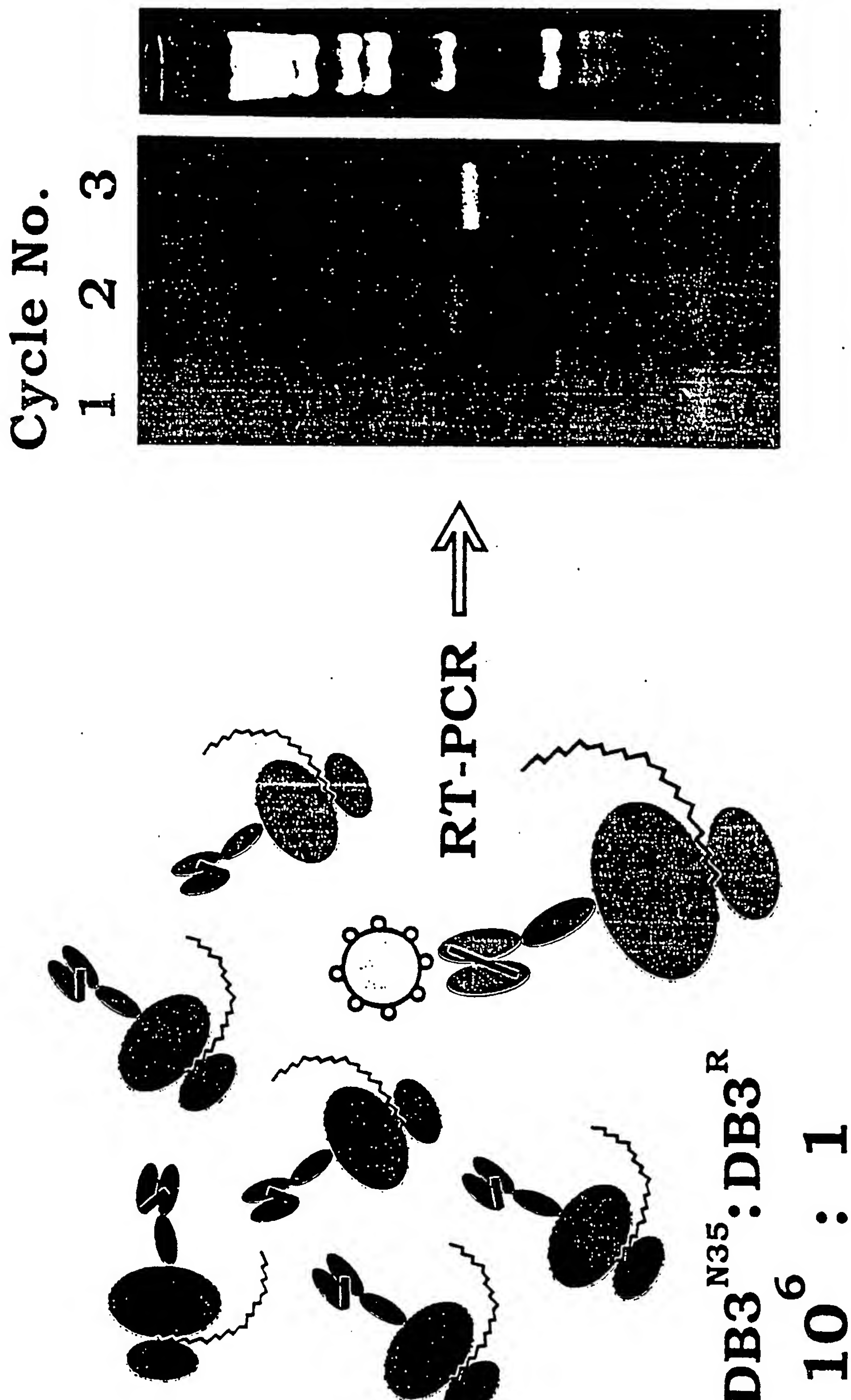
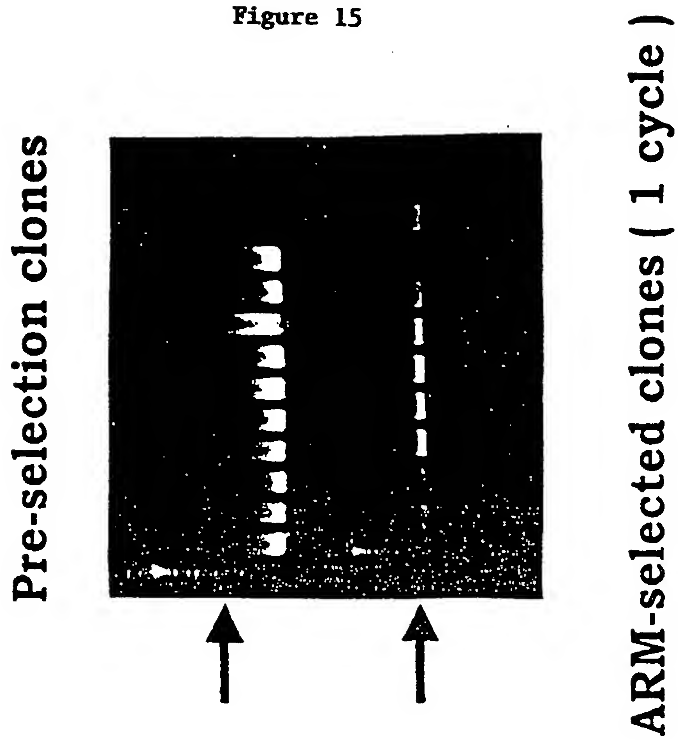


Figure 15



H3 loop of DB3
anti-progesterone
was mutated and
anti-testosterone
selected using
ARM cycle

Cycle 1:
H3 library selected by testosterone beads with free progesterone inhibitor.

Cycle 2:
Testosterone specificity of the selected antibody fragments.

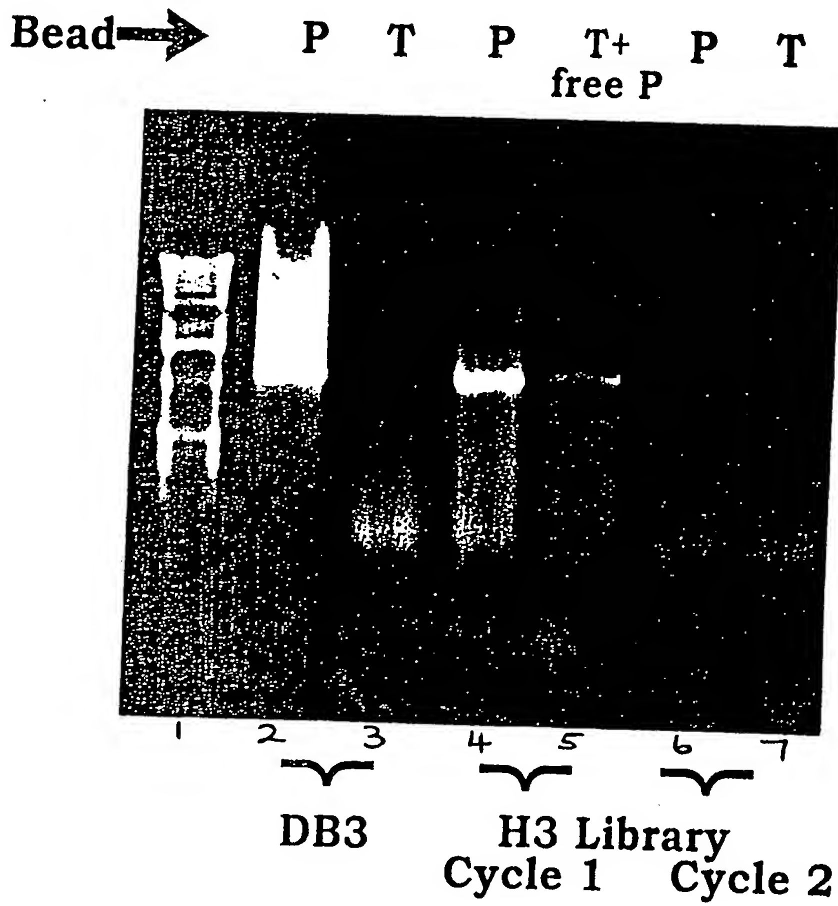


Figure 16

Figure 17

Figure 18

H3 Sequence

DB3R:	GDYVNRYFDVW
Library:	GDT**R***VW
Mut:	GDTRPRSQKVW

P T T+ free P T+ free T

Figure 19

Sequences of human antibody V regions selected by ARM display

VH sequences

Clone	VH gene	CDR H1 31	CDR H2 50	CDR H3 95	
1578/p5	4	SYYWS	WIGRIYTSDSTNYNPSLKS	AITGTAFDI	[SEQ ID 15, 16, 17]
1578/p6	4	SYYWS	WIGRIYTSGSTNYNPSLKG	DSDWNYPFDY	[SEQ ID 15, 18, 19]
1578/p1	1-2	GYYMH	WINPNSGGTNYAQKFQG	YPLLTGDGAFDI	[SEQ ID 20, 21, 22]
1578/p2	1-2	GYYMH	WINPNSGGTNYAQKFQG	DDYEIDWYFGL	[SEQ ID 20, 21, 23]
1578/p9	1-2	GYYMH	WINPNSGGTNYAQKFQG	DLSTEDQAFDI	[SEQ ID 20, 21, 24]
1578/p10	1-2	GYYMH	WINPN??GTNY?QKFQG	DLGNWFDP	[SEQ ID 20, 25, 26]
1578/p11	1-2	GYYMH	WINPNSGGTNYAQKFQG	GSDYGDYEYFQH	[SEQ ID 20, 21, 27]
1578/p14	1-2	GYYMH	WINPNSGGTNYAQKFQG	GSSYGDYEY?QH	[SEQ ID 20, 21, 28]
1578/p16	1-2	GYYMH	WINPNSGGTNYAQKFQG	EYNWFDP	[SEQ ID 20, 21, 29]
1578/t4	1-2	GYYMH	WINPNSGGTNYAQKFQG	QYYDFWSGYYYFDY	[SEQ ID 20, 21, 30]

VL sequences

Clone	VL gene	CDR L1	CDR L2	CDR L3	
1578/p5	1-12	RASQGISRWLA	AGSSLQ		[SEQ ID 21, 32]
1578/p6	1-12	RASQGISSWLA	AASSLQ		[SEQ ID 33, 34]
1578/p1	4-01	SQSVLYSFS?KNYL	ASTRES		[SEQ ID 35, 36]
1578/p2	4-01	SQSVLYSFSNNKNYL	AFTREG		[SEQ ID 37, 38]
1578/t4	4-01	SQSGLYSFNNKNYL			[SEQ ID 39]

p = anti-progesterone
t = anti-testosterone

Figure 19

Sequences of human antibody V regions selected by ARM display

VH sequences

Clone	VH gene	CDR H1	CDR H2	CDR H3
		31	50	95
1578/p5	4	SYYWS	WIGRIYTSDSTNYNPSLKS	AITGTAFDI
1578/p6	4	SYYWS	WIGRIYTSGSTNYNPSLKG	DSDWNYPFDY
1578/p1	1-2	GYYMH	WINPNSGGTNYAQKFQG	YPLLTGDGAFDI
1578/p2	1-2	GYYMH	WINPNSGGTNYAQKFQG	DDYEIDWYFGL
1578/p9	1-2	GYYMH	WINPNSGGTNYAQKFQG	DLSTEDQAFDI
1578/p10	1-2	GYYMH	WINPN??GTNY?QKFQG	DLGNWFDP
1578/p11	1-2	GYYMH	WINPNSGGTNYAQKFQG	GSDYGDYEYFQH
1578/p14	1-2	GYYMH	WINPNSGGTNYAQKFQG	GSSYGDYEY?QH
1578/p16	1-2	GYYMH	WINPNSGGTNYAQKFQG	EYNWFDP
1578/t4	1-2	GYYMH	WINPNSGGTNYAQKFQG	QYYDFWSGYYYFDY

VL sequences

Clone	VL gene	CDR L1	CDR L2	CDR L3
1578/p5	1-12	RASQGISRWLA	AGSSLQ	
1578/p6	1-12	RASQGISSWLA	AASSLQ	
1578/p1	4-01	SQSVLYSFS?KNYL	ASTRES	
1578/p2	4-01	SQSVLYSFSNNKNYL	AFTREG	
1578/t4	4-01	SQSGLYSFNNKNYL		

p = anti-progesterone
t = anti-testosterone

Figure 20

INTERNA‾‾ONAL SEARCH REPORT

.ational Application No
. ᴗT/GB 98/01564

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/10 C12P19/34 C12Q1/68 C12N15/13

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C12Q C07K C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HANES ET AL.: "In vitro selection and evolution of functional proteins by using ribosome display" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, vol. 94, no. 10, 12 May 1997, pages 4937-4942, XP002079230 cited in the application see the whole document ---	13-23, 28-30
X	WO 95 11922 A (AFFYMAX TECHNOLOGIES N.V.) 4 May 1995 cited in the application see the whole document ---	13-23, 28-30
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 October 1998	20/10/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Muller-Thomalla, K

2

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	MATTHEAKIS ET AL.: "Cell-free synthesis of peptide libraires displayed on polysomes" METHODS IN ENZYMOLOGY, vol. 267, 1996, page 195-207 XP002079232 cited in the application see the whole document	13-23, 28-30
X	JP 06 178688 A (MITSUI TOATSU CHEM INC) 28 June 1994 see the whole document	11
X,P	HE ET AL.: "Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for in vitro display and evolution of antibody combining sites" NUCLEIC ACIDS RESEARCH, vol. 25, no. 24, 15 December 1997, pages 5132-5134, XP002079231 see the whole document	1-30
A	HE ET AL.: "Characterization of a progesterone-binding, three-domain antibody fragment (VH/K) expressed in Escherichia coli" IMMUNOLOGY, vol. 84, no. 4, April 1995, pages 662-668, XP002079233 cited in the application see the whole document	11

INTERNATIONAL SEARCH REPORT

Inf on on patent family members

Int ... tional Application No
GB 98/01564

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9511922 A	04-05-1995	AU 8124694 A	22-05-1995
JP 6178688 A	28-06-1994	NONE	

Form PCT/ISA/210 (patent family annex) (July 1992)

PCT

..D INTELLECTUAL PROPERTY ORGANIZATIO.
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [6] :
C12Q 1/68

A2

(11) International Publication Number: WO 99/35293

(43) International Publication Date: 15 July 1999 (15.07.99)

(21) International Application Number: PCT/US99/00666

(22) International Filing Date: 8 January 1999 (08.01.99)

(30) Priority Data:
09/005,222 9 January 1998 (09.01.98) US
09/130,446 6 August 1998 (06.08.98) US

(71) Applicant *(for all designated States except US)*: LYNX THERAPEUTICS, INC. [US/US]; 3832 Bay Center Place, Hayward, CA 94545 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: ALBRECHT, Glen [US/US]; 1113 Oliver Street, Redwood City, CA 94061 (US). BRENNER, Sydney [GB/GB]; 17 B St. Edwards Passage, Cambridge CB2 3PJ (GB). DUBRIDGE, Robert [US/US]; 825 Holly Road, Belmont, CA 94002 (US).

(74) Agents: ASHTON, Nina, M.; Cooley Godward LLP, Five Palo Alto Square, 3000 El Camino Real, Palo Alto, CA 94306-2155 (US) et al.

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: SOLID PHASE SELECTION OF DIFFERENTIALLY EXPRESSED GENES

(57) Abstract

The invention provides a method and materials for monitoring and isolating differentially expressed genes. In accordance with the method of the invention, differently labeled populations of DNAs from sources to be compared are competitively hybridized with reference DNA cloned on solid phase supports, *e.g.* microparticles, to provide a differential expression library which, in the preferred embodiment, may be manipulated by fluorescence-activated cell sorting (FACS). Monitoring the relative signal intensity of the different fluorescent labels on the microparticles permits quantitative analysis of expression levels relative to the reference DNA. The invention also provides a method for identifying and isolating rare genes. Populations of microparticles having relative signal intensities of interest can be isolated by FACS and the attached DNAs identified by sequencing, such as with massively parallel signature sequencing (MPSS), or with conventional DNA sequencing protocols.

SOLID PHASE SELECTION OF DIFFERENTIALLY EXPRESSED GENES

This is a continuation-in-part of co-pending U.S. patent application Ser. No. 09/130,446 filed 6 August 1998, which is a continuation-in-part of co-pending U.S. patent application Ser. No. 09/005,222 filed 9 January 1998, which applications are incorporated by reference.

FIELD OF THE INVENTION

The invention relates generally to methods for identifying differentially expressed genes, and more particularly, to a method of competitively hybridizing differentially expressed DNAs with reference DNA sequences cloned on solid phase supports to provide a differential expression library which can be physically manipulated, *e.g.* by fluorescence-activated flow sorting.

BACKGROUND

The desire to decode the human genome and to understand the genetic basis of disease and a host of other physiological states associated differential gene expression has been a key driving force in the development of improved methods for analyzing and sequencing DNA, Adams et al., Editors, Automated DNA Sequencing and Analysis (Academic Press, New York, 1994). The human genome is estimated to contain about 10^5 genes, 15-30% of which--or about 20-40 megabases--are active in any given tissue. Such large numbers of expressed genes make it difficult to track changes in expression patterns by available techniques, especially in view of the large number of genes that are expressed at relative low levels: It has been estimated that as much as 30% of mRNA consists of many thousands of distinct species each making up less than 0.5% of the total, and typically averaging less than 14 copies per cell, Sambrook et al., Molecular Cloning, Second Edition (Cold Spring Harbor Laboratory Press, New York, 1989). Even substantial changes in expression among such low abundance mRNAs can be difficult to detect in the presence overwhelming quantities of abundant sequences.

A variety of techniques are available for analyzing gene expression that differ widely in convenience, expense, and sensitivity. Commonly used low resolution

techniques include differential display, indexing, subtraction hybridization, and numerous DNA fingerprinting techniques, e.g. Vos et al., Nucleic Acids Research, 23: 4407-4414 (1995); Hubank et al., Nucleic Acids Research, 22: 5640-5648 (1994); Lingo et al., Science, 257: 967-971 (1992); Erlander et al., International patent application PCT/US94/13041; McClelland et al., U.S. patent 5,437,975; Unrau et al., Gene, 145: 163-169 (1994); and the like. Higher resolution techniques include analysis of expressed sequence tags (ESTs), e.g. Adams et al. (cited above); analysis of concatenated fragments of expressed sequences (SAGE), e.g. Velculescu et al., Science, 270: 484-486 (1995); Zhang et al., Science, 276: 1268-1272 (1997); Velculescu et al., Cell, 88: 243-251 (1997); and the use of microarrays of oligonucleotides or polynucleotides for capturing complementary polynucleotides from expressed genes, e.g. Schena et al., Science, 270: 467-469 (1995); DeRisi et al., Science, 278: 680-686 (1997); Chee et al., Science, 274: 610-614 (1996); and the like.

The latter two high resolution techniques have shown promise as potentially robust systems for analyzing gene expression; however, there are still technical issues that need to be addressed with both approaches. In microarray systems, genes to be monitored must be known and isolated beforehand, which means different microarrays, or "DNA chips," have to be manufactured for each specialized use and for every different type of organism or species examined. With respect to microarrays constructed from fluid-delivered cDNAs, a significant degree of variability, e.g. 2-5 fold, exists in the signals generated under the same hybridization conditions, Atlas™ cDNA Expression System Users Manual (Clontech Laboratories, Palo Alto, 1998), and the systems are not readily re-usable. With respect to microarrays of synthetic oligonucleotides, a significant set-up cost for manufacturing such arrays and expensive chip-reading instruments put such systems beyond the financial capability of many potential users. In sequence tag systems, although no special instrumentation is necessary, as an extensive installed base of DNA sequencers may be used, even routine expression analysis requires a significant sequencing effort, e.g. several thousand sequencing reactions or more; the selection of type IIs tag-generating enzymes is limited; and the length (nine nucleotides) of the sequence tag in current protocols severely limits the number of cDNAs that can be uniquely labeled. It can be shown that for organisms expressing large sets of genes, such as mammalian cells, the likelihood of nine-nucleotide tags being distinct for all expressed genes is extremely

low, e.g. Feller, An Introduction to Probability Theory and Its Applications, Second Edition, Vol. I (John Wiley & Sons, New York, 1971).

It is clear from the above that there is a need for a convenient and sensitive technique for analyzing gene expression that permits the analysis of either known or unknown genes from any source. The availability of such a technique would find immediate application not only in medical and scientific research, but also in a host of applied fields, such as crop and livestock development, pest management, drug development, diagnostics, disease management, and the like.

SUMMARY OF THE INVENTION

Accordingly, objects of our invention include, but are not limited to, providing a method for identifying and isolating differentially expressed genes; providing a method of identifying and isolating polynucleotides on the basis of labels that generate different optical signals; providing a method for profiling gene expression of large numbers of genes simultaneously; providing a method of identifying and separating genes in accordance with whether their expression is increased or decrease under any given conditions; providing a method for identifying rare genes; and providing a method for massively parallel signature sequencing of large numbers of genes isolated according to their expression.

Our invention accomplishes these and other objects by providing differently labeled populations of polynucleotides from cell or tissue sources whose gene expression is to be compared. In comparing gene expression, differently labeled polynucleotides of a plurality of populations are competitively hybridized with reference DNA cloned on solid phase supports. Preferably, the solid phase supports are microparticles which, after such competitive hybridization, provide a differential expression library which may be manipulated by fluorescence-activated cell sorting (FACS), or other sorting means responsive to optical signals generated by labeled polynucleotides on the microparticles. Monitoring the relative signal intensity of the different labels on the microparticles permits quantification of the relative expression of particular genes in the different populations.

In one aspect of the invention, populations of microparticles having relative signal intensities of interest are isolated by FACS and the attached polynucleotides are sequenced to determine the identities of the rare or differentially expressed genes.

Preferably, the method of the invention is carried out by the following steps: a) providing a reference population of nucleic acid sequences attached to separate solid phase supports in clonal subpopulations; b) providing a population of polynucleotides of expressed genes from each of the plurality of different cells or tissues, the polynucleotides of expressed genes from different cells or tissues having a different light-generating label; c) competitively hybridizing the populations of polynucleotides of expressed genes from each of the plurality of different cells or tissues with the reference population to form duplexes between the sequences of the reference population and polynucleotides of each of the different cells or tissues such that the polynucleotides are present in duplexes on each of the solid phase supports in ratios directly related to the relative expression of their corresponding genes in the different cells or tissues; and d) detecting a relative optical signal generated by the light-generating labels of the duplexes attached thereto. In further preference, the method includes the step of sorting each solid phase support according to the relative optical signal detected. Preferably, the reference population of nucleic acids is derived from genes of the plurality of different cells or tissues being analyzed. As used herein, the phrase "polynucleotides of expressed genes" is meant to include any RNA produced by transcription, including in particular mRNA, and DNA produced by reverse transcription of any RNA, including in particular cDNA produced by reverse transcription of mRNA.

The present invention overcomes shortcoming in the art by providing compositions, methods, and kits for separating and identifying genes that are differentially expressed without requiring any previous analysis or knowledge of the sequences. The invention also permits differentially regulated genes to be separated from unregulated genes for analysis, thereby eliminating the need to analyze large numbers of unregulated genes in order to obtain information on the genes of interest.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1a and 1b illustrate FACS analysis of microparticles loaded with competitively hybridized DNA strands labeled with two different fluorescent dyes.

Figure 2 is a schematic representation of a flow chamber and detection apparatus for observing a planar array of microparticles loaded with restriction fragments for sequencing.

Figure 3a illustrates a preferred scheme for converting isolated messenger RNA (mRNA) into cDNA and insertion of the cDNA into a tag-containing vector.

Figure 3b illustrates a preferred scheme for amplifying tag-cDNA conjugates out of a vector and loading the amplified conjugates onto microparticles.

Figure 3c illustrates a preferred scheme for isolating sorted cDNAs for cloning and sequencing.

Figure 4a and 4b illustrate alternative procedures for cloning differentially expressed cDNAs isolated by FACS sorting.

Figures 5a-e illustrate flow analysis data of microparticles carrying predetermined ratios of two differently labeled cDNAs.

Figure 6 illustrates flow analysis data of microparticles carrying differently labeled cDNAs from stimulated and unstimulated THP-1 cells.

Figure 7 illustrates flow analysis data of microparticles carrying labeled cDNAs derived from mRNA of low abundance in stimulated THP-1 cells.

Figure 8 illustrates flow analysis data of microparticles carrying labeled cDNAs derived from mRNA of low abundance in human bone marrow.

Figure 9 illustrates flow analysis data of microparticles carrying differently labeled cDNAs from glucose normal and glucose starved muscle tissue.

Figure 10A illustrates an embodiment of the invention for constructing a reference nucleic acid population on microparticles.

Figure 10B illustrates an embodiment for using the reference library of Figure 10A to compare gene expression of two cell populations.

Definitions

"Complement" or "tag complement" as used herein in reference to oligonucleotide tags refers to an oligonucleotide to which a oligonucleotide tag specifically hybridizes to form a perfectly matched duplex or triplex. In embodiments where specific hybridization results in a triplex, the oligonucleotide tag may be selected to be either double stranded or single stranded. Thus, where triplexes are formed, the term "complement" is meant to encompass either a double stranded complement of a single stranded oligonucleotide tag or a single stranded complement of a double stranded oligonucleotide tag.

The term "oligonucleotide" as used herein includes linear oligomers of natural or modified monomers or linkages, including deoxyribonucleosides, ribonucleosides, anomeric forms thereof, peptide nucleic acids (PNAs), and the like, capable of specifically binding to a target polynucleotide by way of a regular pattern of monomer-to-monomer interactions, such as Watson-Crick type of base pairing, base stacking, Hoogsteen or reverse Hoogsteen types of base pairing, or the like. Usually monomers are linked by phosphodiester bonds or analogs thereof to form oligonucleotides ranging in size from a few monomeric units, *e.g.* 3-4, to several tens of monomeric units, *e.g.* 40-60. Whenever an oligonucleotide is represented by a sequence of letters, such as "ATGCCTG," it will be understood that the nucleotides are in 5'→3' order from left to right and that "A" denotes deoxyadenosine, "C" denotes deoxycytidine, "G" denotes deoxyguanosine, and "T" denotes thymidine, unless otherwise noted. Usually oligonucleotides of the invention comprise the four natural nucleotides; however, they may also comprise non-natural nucleotide analogs. It is clear to those skilled in the art when oligonucleotides having natural or non-natural nucleotides may be employed, *e.g.* where processing by enzymes is called for, usually oligonucleotides consisting of natural nucleotides are required.

"Perfectly matched" in reference to a duplex means that the poly- or oligonucleotide strands making up the duplex form a double stranded structure with one other such that every nucleotide in each strand undergoes Watson-Crick basepairing with a nucleotide in the other strand. The term also comprehends the pairing of nucleoside analogs, such as deoxyinosine, nucleosides with 2-aminopurine bases, and the like, that may be employed. In reference to a triplex, the term means that the triplex consists of a perfectly matched duplex and a third strand in which

every nucleotide undergoes Hoogsteen or reverse Hoogsteen association with a basepair of the perfectly matched duplex. Conversely, a "mismatch" in a duplex between a tag and an oligonucleotide means that a pair or triplet of nucleotides in the duplex or triplex fails to undergo Watson-Crick and/or Hoogsteen and/or reverse Hoogsteen bonding.

As used herein, "nucleoside" includes the natural nucleosides, including 2'-deoxy and 2'-hydroxyl forms, *e.g.* as described in Kornberg and Baker, DNA Replication, 2nd Ed. (Freeman, San Francisco, 1992). "Analogs" in reference to nucleosides includes synthetic nucleosides having modified base moieties and/or modified sugar moieties, *e.g.* described by Scheit, Nucleotide Analogs (John Wiley, New York, 1980); Uhlman and Peyman, Chemical Reviews, 90: 543-584 (1990), or the like, with the only proviso that they are capable of specific hybridization. Such analogs include synthetic nucleosides designed to enhance binding properties, reduce complexity, increase specificity, and the like.

As used herein "sequence determination" or "determining a nucleotide sequence" in reference to polynucleotides includes determination of partial as well as full sequence information of the polynucleotide. That is, the term includes sequence comparisons, fingerprinting, and like levels of information about a target polynucleotide, as well as the express identification and ordering of nucleosides, usually each nucleoside, in a target polynucleotide. The term also includes the determination of the identification, ordering, and locations of one, two, or three of the four types of nucleotides within a target polynucleotide. For example, in some embodiments sequence determination may be effected by identifying the ordering and locations of a single type of nucleotide, *e.g.* cytosines, within the target polynucleotide "CATCGC ..." so that its sequence is represented as a binary code, *e.g.* "100101 ... " for "C-(not C)-(not C)-C-(not C)-C ... " and the like.

As used herein, the term "complexity" in reference to a population of polynucleotides means the number of different species of polynucleotide present in the population.

As used herein, the term "relative gene expression" or "relative expression" in reference to a gene refers to the relative abundance of the same gene expression product, usually an mRNA, in different cells or tissue types.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides compositions, methods, and kits for analyzing relative gene expression in a single or plurality of cell and/or tissue types that are of interest. The methods of the invention can be applied to polynucleotides derived from animals, plants, and microorganisms such as fungi, bacteria, mycoplasma, cyanobacteria, algae, and the like. Preferably, the polynucleotides are derived from animals, plants or microorganisms involved in fermentation process, with vertebrates and agricultural plants being most preferred. The plurality usually comprises a pair of cell or tissue types, such as a diseased tissue or cell type and a healthy tissue or cell type, or such as a cell or tissue type being subjected to a stimulus or stress, *e.g.* a change of nutrients, temperature, or the like, and the corresponding cell or tissue type in an unstressed or unstimulated state. In another embodiment, the plurality can comprise a pair of cell or tissue types having homologous genes, such as cells or tissue from different organisms. The plurality may also include more than two cell or tissue types, such as would be required in a comparison of expression patterns of the same cell or tissue over time, *e.g.* liver cells after exposure of an organism to a candidate drug, organ cells of a test animal at different developmental states, and the like. Preferably, the plurality is 2 or 3 cell or tissue types; and more preferably, it is 2 cell or tissue types.

The method of the invention typically comprises providing a reference population of nucleic acid sequences attached to separate solid phase supports in clonal subpopulations, providing at least one population of polynucleotides of expressed genes, hybridizing the population(s) of polynucleotides of expressed genes with the reference nucleic acid population, and detecting, and preferably sorting each solid phase support according to a relative optical signal generated by the duplexes attached thereto.

Figure 10A illustrates an embodiment of the invention for constructing a reference nucleic acid population on microparticles, and Figure 10B illustrates an embodiment for using such a reference library to compare gene expression of two cell populations. Messenger RNA (mRNA) is extracted (1004) from cell populations (1000) and (1002) using conventional protocols to give two populations of polynucleotides (1006) and (1008), respectively. The extraction reactions can be carried out separately or on a mixture of cell types. Preferably, the reactions are

carried out separately so that the relative quantities of mRNA from the two populations can be more readily controlled. Portions of mRNA (1006) and mRNA (1008) are combined (1010) and cDNA library (1012) is constructed in vectors carrying a repertoire of oligonucleotide tags, in accordance with the procedure described in Brenner et al., U.S. patent 5,846,719. Preferably, equal portions of mRNA, *i.e.*, equal molar quantities, are taken from each population of mRNA. A sample of vectors from library (1012) is taken and amplified, *e.g.* by polymerase chain reaction, transfection and cloning, or the like, after which the tag-cDNA conjugates (1014) carried by the vectors are excised or copied (1011) and then isolated. Loaded microparticles are then formed and prepared for use in competitive hybridization as follows (1018). The isolated tag-cDNA conjugates (1014), illustrated with oligonucleotide tags a, b, c, and d, are specifically hybridized to microparticles carrying their tag complements a', b', c', and d' (1016), respectively. The tag-cDNA conjugates are ligated to tag complements so that at least one strand of the double stranded tag-cDNA conjugate is covalently attached to the microparticle. Microparticles carrying tag-cDNA conjugates are separated from those that do not carry tag-cDNA conjugates, preferably using a fluorescence-activated cell sorter (FACS), or like instrument. The non-covalently attached strand is melted off and separated from the microparticles to yield microparticles (1020) carrying a reference nucleic acid population.

As illustrated in Figure 10b, gene expression of cells (1050) may be compared to that of cells (1052) by separately extracting (1054) mRNA (1056) and (1058) from each cell type. After construction of cDNA libraries (1062) and (1064) using conventional protocols, single stranded nucleic acid probes are generated from the respective cDNA populations (1062) and (1064), the probes preferably being labeled with optically distinguishable fluorescent dyes F (1068) and R (1066), *e.g.*, rhodamine and fluorescein. Equal amounts of the labeled polynucleotides are mixed and hybridized (1072) to the complementary strands carried by the microparticles to form duplexes (1074). After the hybridization is complete, microparticles carrying the duplexes thereby formed (1074) can be sorted (1076) in accordance to predetermined criteria, such as fluorescence ratio, fluorescence intensity, and/or the like. In such a manner, subpopulations of interest can be isolated and further analyzed, *e.g.*, those corresponding to up-regulated or down-regulated genes.

For analysis in accordance with the invention, messenger RNA (mRNA) is extracted from the cells or tissues of interest using conventional protocols, as disclosed in, for example, Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edition (Cold Spring Harbor Laboratory, New York). Preferably, the populations of mRNAs to be compared are converted into populations of labeled cDNAs by reverse transcription in the presence of a labeled nucleoside triphosphate using conventional protocols, *e.g.* Schena et al., Science 270: 467-470 (1995); DeRisi et al., Science 278: 680-686 (1997); or the like, prior to hybridization to a reference DNA population.

An important feature of the invention is that the genes whose expression levels change or are different than those of the other cells or tissues being examined may be analyzed separately from those that are not regulated or otherwise altered in response to whatever stress or condition is being studied. As described below, in the preferred embodiment gene products from the cells or tissues of interest are competitively hybridized with a reference population consisting of DNA sequences attached in clonal subpopulations to separate microparticles.. As a result, microparticles carrying labeled gene products in ratios indicating differential expression may be manipulated and analyzed separately from those carrying labeled gene products in ratios indicating no change in expression, *e.g.* "house-keeping" genes, genes encoding structural proteins, or the like.

Another important feature of the invention is that the identity of the nucleic acid being analyzed, *e.g.*, genomic DNA or gene products such as cDNA, mRNA, RNA transcript, or the like, need not be known prior to analysis. After relative expression is determined, cDNAs derived from expressed genes may be identified by direct sequencing on the solid phase support, preferably a microparticle, using a number of different sequencing approaches. For identification, only a portion of the cDNAs need be sequenced. In many cases, the portion may be as small as nine or ten nucleotides, *e.g.* Velculescu et al. (cited above). Preferably, entire subpopulations of differentially expressed genes are sequenced simultaneously using MPSS, or a similar parallel analysis technique. In a preferred embodiment, this is conveniently accomplished by providing a reference population of DNA sequences such that each such sequence is attached to a separate microparticle in a clonal subpopulation. As used herein, the phrase "clonal subpopulation" refers to multiple copies of a single

kind of polynucleotide selected from a population of interest, such as a cDNA library constructed from mRNA extracted from a cell or tissue whose gene expression is being analyzed. Such clonal subpopulations may be formed in a number of ways, including by separate amplification of a poynucleotide and attacment by conventional attachment chemistries, *e.g.*, Hermansen, Bioconjugate Techniques (Academic Press, New York, 1996). As explained more fully below, clonal subpopulations are preferably formed by so-called "solid phase cloning" disclosed in Brenner, U.S. patent 5,604,097 and Brenner et al., U.S. patent 5,846,719, which are incorporated herein by reference. Briefly, such clonal subpopulations are formed by hybridizing an amplified sample of tag-DNA conjugates onto one or more solid phase support(s), *e.g.*, separate and unconnected microparticles, so that individual microparticles, or different regions of a larger support, have attached multiple copies of the same DNA.

The DNA component of the tag-DNA conjugate can be cDNA, genomic DNA, a fragment of cDNA or genomic DNA, or a synthetic DNA, such as, for example, an oligonucleotide. Preferably the tag-DNA conjugate is a cDNA or a fragment of genomic DNA ("gDNA"). The number of copies of a cDNA or gDNA in a clonal subpopulation may vary widely in different embodiments depending on several factors, including the density of tag complements on the solid phase supports, the size and composition of microparticle used, the duration of hybridization reaction, the complexity of the tag repertoire, the concentration of individual tags, the tag-DNA sample size, the labeling means for generating optical signals, the particle sorting means, signal detection system, and the like.

Guidance for making design choices relating to these factors is readily available in the literature on flow cytometry, fluorescence microscopy, molecular biology, hybridization technology, and related disciplines, as represented by the references cited herein. Preferably, the number of copies of a cDNA or a gDNA in a clonal subpopulation is sufficient to permit FACS detection and/or sorting of microparticles, wherein fluorescent signals are generated by one or more fluorescent dye molecules carried by the cDNAs attached to the microparticles. Typically, this number can be as low as a few thousand, *e.g.* 3,000-5,000, when a fluorescent molecule such as fluorescein is used, and as low as several hundred, *e.g.* 800-8000, when a rhodamine dye, such as rhodamine 6G, is used. More preferably, when loaded microparticles are detected and/or sorted by FACS or like instruments, clonal

subpopulations consist of at least 10^4 copies of a cDNA or gDNA; and most preferably, in such embodiments, clonal subpopulations consist of at least 10^5 copies of a cDNA or gDNA.

Labeled cDNAs or RNAs from the cells or tissues to be compared are competitively hybridized to the DNA sequences of the reference DNA population using conventional hybridization conditions, *e.g.* such as disclosed in Schena et al. (cited above); DeRisi et al. (cited above); or Shalon, Ph.D. Thesis entitled "DNA Microarrays," Stanford University (1995). After hybridization, an optical signal is generated by each of the two labeled species of cDNAs or RNAs so that a relative optical signal is determined for each microparticle. Preferably, such optical signals are generated and measured in a fluorescence activated cell sorter, or like instrument, which permits the microparticles to be sorted and accumulated whose relative optical signal fall with a predetermined range of values. The microparticles loaded with cDNAs or RNAs generating relative optical signals in the desired range may be isolated and identified by sequencing, such as with MPSS, as described more fully below.

Preferably, clonal subpopulations of cDNAs or other DNA molecules derived from RNA are attached to microparticles using the processes illustrated in Figures 3a and 3b. First, as illustrated in Figure 3a, mRNA (300) is extracted from a cell or tissue source of interest using conventional techniques and is converted into cDNA (309) with ends appropriate for inserting into vector (316). Preferably, primer (302) having a 5' biotin (305) and poly(dT) region (306) is annealed to mRNA strands (300) so that the first strand of cDNA (309) is synthesized with a reverse transcriptase in the presence of the four deoxyribonucleoside triphosphates. Preferably, 5-methyldeoxycytidine triphosphate is used in place of deoxycytosine triphosphate in the first strand synthesis, so that cDNA (309) is hemi-methylated, except for the region corresponding to primer (302). This allows primer (302) to contain a non-methylated restriction site for releasing the cDNA from a support. The use of biotin in primer (302) is not critical to the invention and other molecular capture techniques, or moieties, can be used, *e.g.* triplex capture, or the like. Region (303) of primer (302) preferably contains a sequence of nucleotides that results in the formation of restriction site r_2 (304) upon synthesis of the second strand of cDNA (309). After isolation by binding the biotinylated cDNAs to streptavidin supports, *e.g.* Dynabeads

M-280 (Dynal, Oslo, Norway), or the like, cDNA (309) is preferably cleaved with a restriction endonuclease which is insensitive to hemimethylation (of the C's) and which recognizes site r_1 (307). Preferably, r_1 is a four-base recognition site, *e.g.* corresponding to Dpn II, or like enzyme, which ensures that substantially all of the cDNAs are cleaved and that the same defined end is produced in all of the cDNAs. After washing, the cDNAs are then cleaved with a restriction endonuclease recognizing r_2, releasing fragment (308) which is purified using standard techniques, *e.g.* ethanol precipitation, polyacrylamide gel electrophoresis, or the like. After resuspending in an appropriate buffer, fragment (308) is directionally ligated into vector (316), which carries tag (310) and a cloning site with ends (312) and (314). Preferably, vector (316) is prepared with a "stuffer" fragment in the cloning site to aid in the isolation of a fully cleaved vector for cloning.

Preparation of the tag-cDNA conjugates is not limited to the method described above and can readily be achieved in a variety of ways using conventional molecular biology techniques. For example, cDNA can be prepared by conventional methods and isolated by gel electrophoresis. This method is less preferred in part because it would bias the size distribution of the reference population. The tag can be attached by ligation of adaptors, by PCR with an oligo dT primer and a random primer, or by RACE technology (Bertling et al. (1993) *PCR Methods Appl.* 3:95-99; Frohman, M.A. (1993) *Methods Enzymol.* 218:340-356; Marathon™ CDNA Amplification Kit, Clontech Laboratories, Inc.). Attachment of the tag by cloning into a vector, as described above, is preferred for several reasons, including the ability to generate large quantities of the reference population (versus RACE, which typically yields only μg quantities), and the ability to check the sequence of the tag.

After formation of a library of tag-cDNA conjugates, a sample of host cells is usually plated to determine the number of recombinants per unit volume of culture medium. The size of sample taken for further processing preferably depends on the size of tag repertoire used in the library construction. As taught by Brenner et al., U.S. patent 5,846,719 and Brenner et al., U.S. patent 5,604,097, a sample preferably includes a number of conjugates equivalent to about one percent the size of the tag repertoire in order to minimize the selection of "doubles," *i.e.* two or more conjugates carrying the same tag and different cDNAs. Thus, for a tag repertoire consisting of a concatenation of eight 4-nucleotide "words" selected from a minimally cross-

hybridizing set of eight words, the size of the repertoire is 8^8, or about 1.7×10^7 tags. Accordingly, with such a repertoire, a sample of about 1.7×10^5 conjugate-containing vectors is preferably selected for amplification and further processing as illustrated in Figure 3b.

Preferably, tag-cDNA conjugates are carried in vector (330) which comprises the following sequence of elements: first primer binding site (332), restriction site r_3 (334), oligonucleotide tag (336), junction (338), cDNA (340), restriction site r_4 (342), and second primer binding site (344). After a sample is taken of the vectors containing tag-cDNA conjugates the following steps are implemented: The tag-cDNA conjugates are preferably amplified from vector (330) by use of biotinylated primer (348) and labeled primer (346) in a conventional polymerase chain reaction (PCR) in the presence of 5-methyldeoxycytidine triphosphate, after which the resulting amplicon is isolated by streptavidin capture. Restriction site r_3 preferably corresponds to a rare-cutting restriction endonuclease, such as Pac I, Not I, Fse I, Pme I, Swa I, or the like, which permits the captured amplicon to be release from a support with minimal probability of cleavage occurring at a site internal to the cDNA of the amplicon. Junction (338) which is illustrated as the sequence:

```
5'... GGGCCC ...
3'... CCCGGG ...
```

causes the DNA polymerase "stripping" reaction to be halted at the G triplet, when an appropriate DNA polymerase is used with dGTP. Briefly, in the "stripping" reaction, the 3'→5' exonuclease activity of a DNA polymerase, preferably T4 DNA polymerase, is used to render the tag of the tag-cDNA conjugate single stranded, as taught by Brenner, U.S. patent 5,604,097; and Kuijper et al., Gene, 112: 147-155 (1992). In the preferred embodiment where sorting is accomplished by formation of duplexes between tags and tag complements, tags of tag-cDNA conjugates are rendered single stranded by first selecting words that contain only three of the four natural nucleotides, and then by preferentially digesting the three nucleotide types from the tag-cDNA conjugate in the 3'→5' direction with the 3'→5' exonuclease activity of a DNA polymerase. In the preferred embodiment, oligonucleotide tags are designed to contain only A's, G's, and T's; thus, tag complements (including that in the

double stranded tag-cDNA conjugate) consist of only A's, C's, and T's. When the released tag-cDNA conjugates are treated with T4 DNA polymerase in the presence of dGTP, the complementary strands of the tags are "stripped" away to the first G. At that point, the incorporation of dG by the DNA polymerase balances the exonuclease activity of the DNA polymerase, effectively halting the "stripping" reaction. From the above description, it is clear that one of ordinary skill could make many alternative design choices for carrying out the same objective, *i.e.* rendering the tags single stranded. Such choices could include selection of different enzymes, different compositions of words making up the tags, and the like.

When the "stripping" reaction is quenched, the result is duplex (356) with single stranded tag (357). After isolation, steps (358) are implemented: the tag-cDNA conjugates are hybridized to tag complements attached to microparticles, a fill-in reaction is carried out to fill any gap between the complementary strand of the tag-cDNA conjugate and the 5' end of tag complement (362) attached to microparticle (360), and the complementary strand of the tag-cDNA conjugate is covalently bonded to the 5' end (363) of tag complement (362) by treating with a ligase. This embodiment requires, of course, that the 5' end of the tag complement be phosphorylated, *e.g.* by a kinase, such as, T4 polynucleotide kinase, or the like. The fill-in reaction is preferably carried out because the "stripping" reaction does not always halt at the first G. Preferably, the fill-in reaction uses a DNA polymerase lacking 5'$\rightarrow$3' exonuclease activity and strand displacement activity, such as T4 DNA polymerase. Also preferably, all four dNTPs are used in the fill-in reaction, in case the "stripping" extended beyond the G triplet.

As explained further below, the tag-cDNA conjugates are hybridized to the full repertoire of tag complements. That is, among the population of microparticles, there are microparticles having every tag sequence of the entire repertoire. Thus, the tag-cDNA conjugates will hybridize to tag complements on only about one percent of the microparticles. Microparticles to which tag-cDNA have been hybridized are referred to herein as "loaded microparticles." For greater efficiency, loaded microparticles are preferably separated from unloaded microparticles for further processing. Such separation is conveniently accomplished by use of a fluorescence-activated cell sorter (FACS), or similar instrument that permits rapid manipulation and sorting of large numbers of individual microparticles. In the embodiment

illustrated in Figure 3b, a fluorescent label, *e.g.* FAM (a fluorescein derivative, Haugland, Handbook of Fluorescent Probes and Research Chemicals, Sixth Edition, (Molecular Probes, Eugene, OR, 1996)) is attached by way of primer (346).

The tag-cDNA can be attached to the tag complement on the microparticles by a procedure omitting or modifying many of the steps discussed above. For example, instead of amplifying the tag-cDNA from vector (330), the tag-cDNA can be cleaved from the vector by restriction digest, stripped, and ligated directly to the tag complement on the microparticles. This procedure omits (1) labeling the tag-cDNA with biotin and FAM, (2) amplifying the tag-cDNA, and (3) isolating the amplicon by streptavidin capture. If desired, loaded microparticles can be isolated by hybridizing with a FAM-labeled primer.

As shown in Figure 3c, after FACS, or like sorting (380), loaded microparticles (360) are isolated, treated to remove label (345), and treated to melt off the non-covalently attached strand. Label (345) is removed or inactivated so that it does not interfer with the labels of the competitively hybridized strands. Preferably, the tag-cDNA conjugates are treated with a restriction endonuclease recognizing site r_1 (342) which cleaves the tag-cDNA conjugates adjacent to primer binding site (344), thereby removing label (345) carried by the "bottom" strand, *i.e.* the strand have its 5' end distal to the microparticle. Preferably, this cleavage results in microparticle (360) with double stranded tag-cDNA conjugate (384) having protruding strand (385). 3'-labeled adaptor (386) is then annealed and ligated to protruding strand (385), after which the loaded microparticles are re-sorted by means of the 3'-label and the strand carrying the 3'-label is melted off to leave a covalently attached single strand of the cDNA (392) ready to accept denatured cDNAs or mRNAs from differentially expressed genes. Preferably, the 3'-labeled strand is melted off with sodium hydroxide treatment, or treatment with like reagent.

Clonal subpopulations of gDNAs can be attached to microparticles in a similar manner. First, genomic DNA is isolated from a cell or tissue source of interest using conventional techniques and is cleaved with at least one restriction endonuclease, which preferably cleaves at a four-base recognition, such as, for example, Dpn II, Sau3A I, Aci I, Alu I, Bfa I, BstU I, Hae III, Hha I, HinPl I, Hpa II, Mbo I, Mse I, Msp I, Nla III, Rsa I, Taq^{α} I, Tsp 509 I, and the like. Preferably, the cleaved fragment has an overhang of at least one base. Alternatively, genomic DNA fragments can be

prepared by shearing or sonicating the isolated genomic DNA. The tag can then be linked to the gDNA in a number of ways, including random primed PCR with primers containing the tag sequence or cloning into a vector containing a tag in a manner similar to that described above for a cDNA reference population. A label such as FAM can be attached in order to monitor the loading of the microparticles. In some instances, directional attachment onto the microparticles can be achieved by amplifying the gDNA with a primer having a consensus sequence, such as, for example, the TATA box, or a sequence complementary to a consensus sequence. When using a gDNA reference population for evaluating gene expression, it may be desirable to reduce noncoding sequence and introns in the gDNA library. For example, a large gDNA library of about 60 x 10^6 microparticles can be reduced to about 30,000-40,000 by culling, using cDNA pools as a probe.

Oligonucleotide Tags for Identification and Solid Phase Cloning

An important feature of the invention is the use of oligonucleotide tags which are members of a minimally cross-hybridizing set of oligonucleotides to construct reference DNA populations attached to solid phase supports, preferably microparticles. The sequences of oligonucleotides of a minimally cross-hybridizing set differ from the sequences of every other member of the same set by at least two nucleotides. Thus, each member of such a set cannot form a duplex (or triplex) with the complement of any other member with less than two mismatches. Complements of oligonucleotide tags, referred to herein as "tag complements," may comprise natural nucleotides or non-natural nucleotide analogs. When oligonucleotide tags are used for sorting, as is the case for constructing a reference DNA population, tag complements are preferably attached to solid phase supports. Oligonucleotide tags when used with their corresponding tag complements provide a means of enhancing specificity of hybridization for sorting, tracking, or labeling molecules, especially polynucleotides, such as cDNAs or mRNAs derived from expressed genes.

Minimally cross-hybridizing sets of oligonucleotide tags and tag complements may be synthesized either combinatorially or individually depending on the size of the set desired and the degree to which cross-hybridization is sought to be minimized (or stated another way, the degree to which specificity is sought to be enhanced). For example, a minimally cross-hybridizing set may consist of a set of individually

synthesized 10-mer sequences that differ from each other by at least 4 nucleotides, such set having a maximum size of 332, when constructed as disclosed in Brenner et al., U.S. patent 5,604,097. Alternatively, a minimally cross-hybridizing set of oligonucleotide tags may also be assembled combinatorially from subunits which themselves are selected from a minimally cross-hybridizing set. For example, a set of minimally cross-hybridizing 12-mers differing from one another by at least three nucleotides may be synthesized by assembling 3 subunits selected from a set of minimally cross-hybridizing 4-mers that each differ from one another by three nucleotides. Such an embodiment gives a maximally sized set of 9^3, or 729, 12-mers.

When synthesized combinatorially, an oligonucleotide tag can be randomized at individual positions along its length. Preferably, however, the oligonucleotide tag consists of a plurality of subunits, each subunit consisting of an oligonucleotide of 3 to 9 nucleotides in length wherein each subunit is selected from the same minimally cross-hybridizing set. In such embodiments, the number of oligonucleotide tags available depends on the number of subunits per tag and on the length of the subunits. An oligonucleotide tag can also consist of a plurality of subunits with additional nucleotides on either terminus of the oligonucleotide. The additional nucleotides can be random and/or can comprise a restriction site. Such a structure ensures the instability of a duplex or triplex having a mismatch at a terminus of the oligonucleotide. Preferably, the oligonucleotide comprises a recognition site for a rare-cutting restriction endonuclease on at least one end. In a preferred embodiment, the oligonucleotide comprises an AT-rich restriction site, such as a Pac I site, on one end. A Bsp120 site is a preferred site on the other end.

Complements of oligonucleotide tags attached to one or more solid phase supports are used to sort polynucleotides from a mixture of polynucleotides each containing a tag. Such tag complements are synthesized on the surface of a solid phase support, such as a bead, preferably microscopic, or a specific location on an array of synthesis locations on a single support, such that populations of identical, or substantially identical, sequences are produced in specific regions. That is, the surface of each support, in the case of a bead, or of each region, in the case of an array, is derivatized by copies of only one type of tag complement having a particular sequence. The population of such beads or regions contains a repertoire of tag complements each with distinct sequences. As used herein in reference to

oligonucleotide tags and tag complements, the term "repertoire" means the total number of different oligonucleotide tags or tag complements that are employed for solid phase cloning (sorting) or identification. A repertoire may consist of a set of minimally cross-hybridizing set of oligonucleotides that are individually synthesized, or it may consist of a concatenation of oligonucleotides each selected from the same set of minimally cross-hybridizing oligonucleotides. In the latter case, the repertoire is preferably synthesized combinatorially.

Preferably, tag complements are synthesized combinatorially on microparticles, so that each microparticle has attached many copies of the same tag complement. A wide variety of microparticle supports may be used with the invention, including microparticles made of controlled pore glass (CPG), highly cross-linked polystyrene, acrylic copolymers, cellulose, nylon, dextran, latex, polyacrolein, and the like, disclosed in the following exemplary references: Meth. Enzymol., Section A, pages 11-147, vol. 44 (Academic Press, New York, 1976); U.S. patents 4,678,814; 4,413,070; and 4,046;720; and Pon, Chapter 19, in Agrawal, editor, Methods in Molecular Biology, Vol. 20, (Humana Press, Totowa, NJ, 1993). Microparticle supports further include commercially available nucleoside-derivatized CPG and polystyrene beads (*e.g.* available from PE Applied Biosystems, Foster City, CA); derivatized magnetic beads; polystyrene grafted with polyethylene glycol (*e.g.*, TentaGelTM, Rapp Polymere, Tubingen Germany); and the like. Microparticles may also consist of dendrimeric structures, such as disclosed by Nilsen et al., U.S. patent 5,175,270. Generally, the size and shape of a microparticle is not critical; however, microparticles in the size range of a few, *e.g.* 1-2, to several hundred, *e.g.* 200-1000 μm diameter are preferable, as they facilitate the construction and manipulation of large repertoires of oligonucleotide tags with minimal reagent and sample usage. Preferably, glycidal methacrylate (GMA) beads available from Bangs Laboratories (Carmel, IN) are used as microparticles in the invention. Such microparticles are useful in a variety of sizes and are available with a variety of linkage groups for synthesizing tags and/or tag complements. More preferably, 5 μm diameter GMA beads are employed.

In a preferred embodiment, polynucleotides to be sorted, or cloned onto a solid phase support, each have an oligonucleotide tag attached, such that different polynucleotides have different tags. This condition is achieved by employing a

repertoire of tags substantially greater than the population of polynucleotides and by taking a sufficiently small sample of tagged polynucleotides from the full ensemble of tagged polynucleotides. After such sampling, when the populations of supports and polynucleotides are mixed under conditions which permit specific hybridization of the oligonucleotide tags with their respective complements, identical polynucleotides sort onto particular beads or regions. Of course, the sampled tag-polynucleotide conjugates are preferably amplified, *e.g.* by polymerase chain reaction, cloning in a plasmid, RNA transcription, or the like, to provide sufficient material for subsequent analysis.

Oligonucleotide tags are employed for two different purposes in certain embodiments of the invention: Oligonucleotide tags are employed to implement solid phase cloning, as described in Brenner, U.S. patent 5,604,097; and International patent application PCT/US96/09513, wherein large numbers of polynucleotides, *e.g.* several thousand to several hundred thousand, are sorted from a mixture into clonal subpopulations of identical polynucleotides on one or more solid phase supports for analysis, and they are employed to deliver (or accept) labels to identify polynucleotides, such as encoded adaptors, that number in the range of a few tens to a few thousand, *e.g.* as disclosed in Albrecht et al., International patent application PCT/US97/09472. For the former use, large numbers, or repertoires, of tags are typically required, and therefore synthesis of individual oligonucleotide tags is difficult. In these embodiments, combinatorial synthesis of the tags is preferred. On the other hand, where extremely large repertoires of tags are not required--such as for delivering labels to a plurality of kinds or subpopulations of polynucleotides in the range of 2 to a few tens, *e.g.* encoded adaptors, oligonucleotide tags of a minimally cross-hybridizing set may be separately synthesized, as well as synthesized combinatorially.

Sets containing several hundred to several thousands, or even several tens of thousands, of oligonucleotides may be synthesized directly by a variety of parallel synthesis approaches, *e.g.* as disclosed in Frank et al., U.S. patent 4,689,405; Frank et al., Nucleic Acids Research, 11: 4365-4377 (1983); Matson et al., Anal. Biochem., 224: 110-116 (1995); Fodor et al., International application PCT/US93/04145; Pease et al., Proc. Natl. Acad. Sci., 91: 5022-5026 (1994); Southern et al., J. Biotechnology,

35: 217-227 (1994), Brennan, International application PCT/US94/05896; Lashkari et al., Proc. Natl. Acad. Sci., 92: 7912-7915 (1995); or the like.

Preferably, tag complements in mixtures, whether synthesized combinatorially or individually, are selected to have similar duplex or triplex stabilities to one another so that perfectly matched hybrids have similar or substantially identical melting temperatures. This permits mis-matched tag complements to be more readily distinguished from perfectly matched tag complements in the hybridization steps, *e.g.* by washing under stringent conditions. For combinatorially synthesized tag complements, minimally cross-hybridizing sets may be constructed from subunits that make approximately equivalent contributions to duplex stability as every other subunit in the set. Guidance for carrying out such selections is provided by published techniques for selecting optimal PCR primers and calculating duplex stabilities, *e.g.* Rychlik et al., Nucleic Acids Research, 17: 8543-8551 (1989) and 18: 6409-6412 (1990); Breslauer et al., Proc. Natl. Acad. Sci., 83: 3746-3750 (1986); Wetmur, Crit. Rev. Biochem. Mol. Biol., 26: 227-259 (1991); and the like. A minimally cross-hybridizing set of oligonucleotides can be screened by additional criteria, such as GC-content, distribution of mismatches, theoretical melting temperature, and the like, to form a subset which is also a minimally cross-hybridizing set.

The oligonucleotide tags of the invention and their complements are conveniently synthesized on an automated DNA synthesizer, *e.g.* an Applied Biosystems, Inc. (Foster City, California) model 392 or 394 DNA/RNA Synthesizer, using standard chemistries, such as phosphoramidite chemistry, *e.g.* disclosed in the following references: Beaucage and Iyer, Tetrahedron, 48: 2223-2311 (1992); Molko et al., U.S. patent 4,980,460; Koster et al., U.S. patent 4,725,677; Caruthers et al., U.S. patents 4,415,732; 4,458,066; and 4,973,679; and the like.

Oligonucleotide tags for sorting may range in length from 12 to 60 nucleotides or basepairs. Preferably, oligonucleotide tags range in length from 18 to 40 nucleotides or basepairs. More preferably, oligonucleotide tags range in length from 25 to 40 nucleotides or basepairs. In terms of preferred and more preferred numbers

of subunits, these ranges may be expressed as follows:

Numbers of Subunits in Tags in Preferred Embodiments

Monomers in Subunit	Nucleotides in Oligonucleotide Tag (12-60)	(18-40)	(25-40)
3	4-20 subunits	6-13 subunits	8-13 subunits
4	3-15 subunits	4-10 subunits	6-10 subunits
5	2-12 subunits	3-8 subunits	5-8 subunits
6	2-10 subunits	3-6 subunits	4-6 subunits

Most preferably, oligonucleotide tags for sorting are single stranded and specific hybridization occurs via Watson-Crick pairing with a tag complement.

Preferably, repertoires of single stranded oligonucleotide tags for sorting contain at least 100 members; more preferably, repertoires of such tags contain at least 1000 members; and most preferably, repertoires of such tags contain at least 10,000 members.

Preferably, the length of single stranded tag complements for delivering labels is between 8 and 20. More preferably, the length is between 9 and 15.

In embodiments where specific hybridization occurs via triplex formation, coding of tag sequences follows the same principles as for duplex-forming tags; however, there are further constraints on the selection of subunit sequences. Generally, third strand association via Hoogsteen type of binding is most stable along homopyrimidine-homopurine tracks in a double stranded target. Usually, base triplets form in T-A*T or C-G*C motifs (where "-" indicates Watson-Crick pairing and "*" indicates Hoogsteen type of binding); however, other motifs are also possible. For example, Hoogsteen base pairing permits parallel and antiparallel orientations between the third strand (the Hoogsteen strand) and the purine-rich strand of the duplex to which the third strand binds, depending on conditions and the composition of the strands. There is extensive guidance in the literature for selecting appropriate sequences, orientation, conditions, nucleoside type (*e.g.* whether ribose or deoxyribose nucleosides are employed), base modifications (*e.g.* methylated cytosine, and the like) in order to maximize, or otherwise regulate, triplex stability as desired in particular embodiments. Conditions for annealing single-stranded or duplex tags to

their single-stranded or duplex complements are well known, *e.g.* Ji et al., Anal. Chem. 65: 1323-1328 (1993); Cantor et al., U.S. patent 5,482,836; and the like. Use of triplex tags in sorting has the advantage of not requiring a "stripping" reaction with polymerase to expose the tag for annealing to its complement.

An exemplary tag library for sorting is shown below (SEQ ID NO: 1).

```
        Left Primer                         Bsp 120I

 5'-AGAATTCGGGCCTTAATTAA                       ↓

 5'-AGAATTCGGGCCTTAATTAA- [4(A,G,T)8] -GGGCCC-
    TCTTAAGCCCGGAATTAATT- [4(T,C,A)8] -CCCGGG-

      ↑             ↑

    Eco RI        Pac  I


                              Bbs I              Bam HI
                                ↓                  ↓

                       -GCATAAGTCTTCXXX ... XXXGGATCCGAGTGAT -3'
                       -CGTATTCAGAAGXXX ... XXXCCTAGGCTCACTA

                                              XXXXXCCTAGGCTCACT
                                              A-5'
                                                 Right Primer
```

Formula I

The flanking regions of the oligonucleotide tag may be engineered to contain restriction sites, as exemplified above, for convenient insertion into and excision from cloning vectors. Optionally, the right or left primers may be synthesized with a biotin attached (using conventional reagents, *e.g.* available from Clontech Laboratories, Palo Alto, CA) to facilitate purification after amplification and/or cleavage. Preferably, for making tag-fragment conjugates, the above library is inserted into a conventional cloning vector, such a pUC19, or the like. Optionally, the vector containing the tag library may contain a "stuffer" region, "XXX ... XXX," which facilitates isolation of fragments fully digested with, for example, Bam HI and Bbs I.

An important aspect of the invention is the sorting and attachment of populations of DNA sequences, *e.g.* from a cDNA library, to microparticles or to

separate regions on a solid phase support such that each microparticle or region has substantially only one kind of sequence attached; that is, such that the DNA sequences are present in clonal subpopulations. This objective is accomplished by insuring that substantially all different DNA sequences have different tags attached. This condition, in turn, is brought about by taking only a sample of the full ensemble of tag-DNA sequence conjugates for analysis. (It is acceptable that identical DNA sequences have different tags, as it merely results in the same DNA sequence being operated on or analyzed twice.) Such sampling can be carried out either overtly--for example, by taking a small volume from a larger mixture--after the tags have been attached to the DNA sequences; it can be carried out inherently as a secondary effect of the techniques used to process the DNA sequences and tags; or sampling can be carried out both overtly and as an inherent part of processing steps.

If a sample of n tag-DNA sequence conjugates are randomly drawn from a reaction mixture--as could be effected by taking a sample volume, the probability of drawing conjugates having the same tag is described by the Poisson distribution, $P(r)=e^{-\lambda}(\lambda)^r/_r$, where r is the number of conjugates having the same tag and $\lambda=np$, where p is the probability of a given tag being selected. If $n=10^6$ and $p=1/(1.67 \times 10^7)$ (for example, if eight 4-base words described in Brenner et al. were employed as tags), then $\lambda=.0149$ and $P(2)=1.13 \times 10^{-4}$. Thus, a sample of one million molecules gives rise to an expected number of doubles well within the preferred range. Such a sample is readily obtained by serial dilutions of a mixture containing tag-fragment conjugates.

As used herein, the term "substantially all" in reference to attaching tags to molecules, especially polynucleotides, is meant to reflect the statistical nature of the sampling procedure employed to obtain a population of tag-molecule conjugates essentially free of doubles. Preferably, at least ninety-five percent of the DNA sequences have unique tags attached.

Preferably, DNA sequences are conjugated to oligonucleotide tags by inserting the sequences into a conventional cloning vector carrying a tag library. For example, cDNAs may be constructed having a Bsp 120 I site at their 5' ends and after digestion with Bsp 120 I and another enzyme such as Sau 3A or Dpn II may be directionally inserted into a pUC19 carrying the tags of Formula I to form a tag-cDNA library, which includes every possible tag-cDNA pairing. A sample is taken from this library

for amplification and sorting. Sampling may be accomplished by serial dilutions of the library, or by simply picking plasmid-containing bacterial hosts from colonies. After amplification, the tag-cDNA conjugates may be excised from the plasmid.

After the oligonucleotide tags are prepared for specific hybridization, *e.g.* by rendering them single stranded as described above, the polynucleotides are mixed with microparticles containing the complementary sequences of the tags under conditions that favor the formation of perfectly matched duplexes between the tags and their complements. There is extensive guidance in the literature for creating these conditions. Exemplary references providing such guidance include Wetmur, Critical Reviews in Biochemistry and Molecular Biology, 26: 227-259 (1991); Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edition (Cold Spring Harbor Laboratory, New York, 1989); and the like. Preferably, the hybridization conditions are sufficiently stringent so that only perfectly matched sequences form stable duplexes. Under such conditions the polynucleotides specifically hybridized through their tags may be ligated to the complementary sequences attached to the microparticles. Finally, the microparticles are washed to remove polynucleotides with unligated and/or mismatched tags.

Specificity of the hybridizations of tag to their complements may be increased by taking a sufficiently small sample so that both a high percentage of tags in the sample are unique and the nearest neighbors of substantially all the tags in a sample differ by at least two words. This latter condition may be met by taking a sample that contains a number of tag-polynucleotide conjugates that is about 0.1 percent or less of the size of the repertoire being employed. For example, if tags are constructed with eight words a repertoire of 8^8, or about 1.67×10^7, tags and tag complements are produced. In a library of tag-DNA sequence conjugates as described above, a 0.1 percent sample means that about 16,700 different tags are present. If this were loaded directly onto a repertoire-equivalent of microparticles, or in this example a sample of 1.67×10^7 microparticles, then only a sparse subset of the sampled microparticles would be loaded. Preferably, loaded microparticles may be separated from unloaded microparticles by a fluorescence activated cell sorting (FACS) instrument using conventional protocols after DNA sequences have been fluorescently labeled and denatured. After loading and FACS sorting, the label may be cleaved prior use or other analysis of the attached DNA sequences.

A reference DNA population may consist of any set of DNA sequences whose frequencies in different test populations is sought to be compared. Preferably, a reference DNA population for use in the analysis of gene expression in a plurality of cells or tissues is constructed by generating a cDNA library from each of the cells or tissues whose gene expression is being compared. This may be accomplished either by pooling the mRNA extracted from the various cells and/or tissues, or it may be accomplished by pooling the cDNAs of separately constructed cDNA libraries. Alternatively, a reference DNA population may be constructed from genomic DNA. The objective is to obtain a set of DNA sequences that will include all of the sequences that could possibly be expressed in any of the cells or tissues being analyzed. Once the DNA sequences making up a reference DNA population are obtained, they must be conjugated with oligonucleotide tags for solid phase cloning. Preferably, the DNA sequences are prepared so that they can be inserted into a vector carrying an appropriate tag repertoire, as described above, to form a library of tag-DNA sequence conjugates. A sample of conjugates is taken from this library, amplified, and loaded onto microparticles. It is important that the sample be large enough so that there is a high probability that all of the different types of DNA sequences are represented on the loaded microparticles. For example, if among a plurality of cells being compared a total of about 25,000 genes are expressed, then a sample of about five-fold this number, or about 125,000 tag-DNA sequence conjugates, should be taken to ensure that all possible DNA sequences will be represented among the loaded microparticles with about a 99% probability, *e.g.* Sambrook et al. (cited above).

In another embodiment, the reference population can comprise a set of polynucleotides encoding a specific set or sets of proteins selected from the group consisting of cell cycle proteins, signal transduction pathway proteins, oncogene gene products, tumor suppressors, kinases, phosphatases, transcription factors, growth factor receptors, growth factors, extracellular matrix proteins, proteases, cytoskeletal proteins, membrane receptors, Rb pathway proteins, p53 pathway proteins, proteins involved in metabolism, proteins involved in cellular responses to stress, cytokines, proteins involved in DNA damage and repair, and proteins involved in apoptosis. Such polynucleotides are typically attached to the solid phase supports through oligonucleotides having a unique sequence per solid support, but such polynucleotides

can also be attached to the solid phase supports through an oligonucleotide with a sequence common for each solid phase support, such as, for example a polyadenylated oligonucleotide.

Preferably, after the tag-DNA sequence conjugates are sampled, they are amplified by PCR using a fluorescently labeled primer to provide sufficient material to load onto the tag complements of the microparticles and to provide a means for distinguishing loaded from unloaded microparticles, as disclosed in Brenner et al. (cited above). Preferably, the PCR primer also contains a sequence which allows the generation of a restriction site of a rare-cutting restriction endonuclease, such as Pac I, in the double stranded product so that the fluorescent label may be cleave from the end of the cDNA prior to the competitive hybridization of labeled DNA strands derived from cells or tissue being studied. After such loading, the specifically hybridized tag-DNA sequence conjugates are ligated to the tag complements and the loaded microparticles are separated from the unloaded microparticles by FACS. The fluorescent label is cleaved from the DNA strands of the loaded microparticles and the non-covalently attached strand is removed by denaturing with heat, formamide, NaOH, and/or with like means, using conventional protocols. The microparticles are then ready for competitive hybridization.

Competitive Hybridization and Light-Generating Labels

Gene expression products, *e.g.* mRNA or cDNA, from the cells and/or tissues being analyzed are isolated. The expression products are labeled so as to distinguish the source. Preferably, the products from each source comprise a label different from the label comprised by the products of any other source, *e.g.*, each having a unique and distinguishable emission frequency. Alternatively, the product of one source can be left unlabeled. The expression products can be labeled by conventional techniques, *e.g.* DeRisi et al. (cited above), or the like. Preferably, a light-generating label is incorporated into cDNAs reverse transcribed from the extracted mRNA, or an oligonucleotide tag is attached for providing a labeled tag complement for identification. A large number of light-generating labels are available, including fluorescent, colorimetric, chemiluminescent, and electroluminescent labels. Generally, such labels produce an optical signal which may comprise an absorption frequency, an emission frequency, an intensity, a signal lifetime, or a combination of

such characteristics. Preferably, fluorescent labels are employed, either by direct incorporation of fluorescently labeled nucleoside triphosphates or by indirect application by incorporation of a capture moiety, such as biotinylated nucleoside triphosphates or an oligonucleotide tag, followed by complexing with a moiety capable of generating a fluorescent signal, such as a streptavidin-fluorescent dye conjugate or a fluorescently labeled tag complement. Preferably, the optical signal detected from a fluorescent label is an intensity at one or more characteristic emission frequencies. Selection of fluorescent dyes and means for attaching or incorporating them into DNA strands is well known, *e.g.* DeRisi et al. (cited above), Matthews et al., Anal. Biochem., Vol 169, pgs. 1-25 (1988); Haugland, Handbook of Fluorescent Probes and Research Chemicals (Molecular Probes, Inc., Eugene, 1992); Keller and Manak, DNA Probes, 2nd Edition (Stockton Press, New York, 1993); and Eckstein, editor, Oligonucleotides and Analogues: A Practical Approach (IRL Press, Oxford, 1991); Wetmur, Critical Reviews in Biochemistry and Molecular Biology, 26: 227-259 (1991); Ju et al., Proc. Natl. Acad. Sci., 92: 4347-4351 (1995) and Ju et al., Nature Medicine, 2: 246-249 (1996); and the like.

Preferably, light-generating labels are selected so that their respective optical signals can be related to the quantity of labeled DNA strands present and so that the optical signals generated by different light-generating labels can be compared. Measurement of the emission intensities of fluorescent labels is the preferred means of meeting this design objective. For a given selection of fluorescent dyes, relating their emission intensities to the respective quantities of labeled DNA strands requires consideration of several factors, including fluorescent emission maxima of the different dyes, quantum yields, emission bandwidths, absorption maxima, absorption bandwidths, nature of excitation light source(s), and the like. Guidance for making fluorescent intensity measurements and for relating them to quantities of analytes is available in the literature relating to chemical and molecular analysis, *e.g.* Guilbault, editor, Practical Fluorescence, Second Edition (Marcel Dekker, New York, 1990); Pesce et al., editors, Fluorescence Spectroscopy (Marcel Dekker, New York, 1971); White et al., Fluorescence Analysis: A Practical Approach (Marcel Dekker, New York, 1970); and the like. As used herein, the term "relative optical signal" means a ratio of signals from different light-generating labels that can be related to a ratio of differently labeled DNA strands of identical, or substantially identical, sequences that

form duplexes with a complementary reference DNA strand. Preferably, a relative optical signal is a ratio of fluorescence intensities of two or more different fluorescent dyes.

Competitive hybridization between the labeled DNA strands derived from the plurality of cells or tissues is carried out by applying equal quantities of the labeled DNA strands from each such source to the microparticles loaded with the reference DNA population in a conventional hybridization reaction. The particular amounts of labeled DNA strands added to the competitive hybridization reaction vary widely depending on the embodiment of the invention. Factors influencing the selection of such amounts include the quantity of microparticles used, the type of microparticles used, the loading of reference DNA strands on the microparticles, the complexity of the populations of labeled DNA strands, and the like. Hybridization is competitive in that differently labeled DNA strands with identical, or substantially identical, sequences compete to hybridize to the same complementary reference DNA strands. The competitive hybridization conditions are selected so that the proportion of labeled DNA strands forming duplexes with complementary reference DNA strands reflects, and preferably is directly proportional to, the amount of that DNA strand in its population in comparison with the amount of the competing DNA strands of identical sequence in their respective populations. Thus, if a first and second differently labeled DNA strands with identical sequence are competing for hybridization with a complementary reference DNA strand such that the first labeled DNA strand is at a concentration of 1 ng/μl and the second labeled DNA strand is at a concentration of 2 ng/μl, then at equilibrium it is expected that one third of the duplexes formed with the reference DNA would include first labeled DNA strands and two thirds of the duplexes would include second labeled DNA strands. Guidance for selecting hybridization conditions is provided in many references, including Keller and Manak, (cited above); Wetmur, (cited above); Hames et al., editors, Nucleic Acid Hybridization: A Practical Approach (IRL Press, Oxford, 1985); and the like.

Another aspect of the invention is a kit for analyzing differentially expressed genes, comprising a mixture of microparticles, each microparticle having a population of identical single stranded nucleic acid molecules attached thereto, the single stranded nucleic acid molecules being different on each microparticle and comprising a polynucleotide derived from an mRNA of at least one cell or tissue source.

Preferably, each of said nucleic acid molecules further comprises an oligonucleotide tag in juxtaposition with said polynucleotide and positioned between said microparticle and said polynucleotide. The kit can further comprise a population of cDNA molecules from at least one of said cell or tissue sources, reagents for labeling the cDNA populations, reagents for performing competitive hybridization, and the like. If desired, the cDNA molecules in the kit are provided in fluorescently labeled form. The kit can contain additional components for performing competitive hybridization, such as, for example, hybridization buffers, PCR buffers and standards, and the like. The kit can further comprise at least one container or several containers for each of the components and can comprise printed instructions for use in analyzing differentially expressed genes.

The invention also provides a kit for preparing a reference population, comprising a plurality of microparticles having oligonucleotide tag complements attached thereto, the oligonucleotide tag complement sequence being different on each microparticle. The kit can further comprise a plurality of vectors comprising a library of tags having sequences complementary to the tag complements. The kit can further comprise a population of polynucleotides from at least one cell or tissue source, preferably cDNAs. When a population of polynucleotides is included, preferably the population of polynucleotides is contained in a container separate from said plurality of microparticles. The kit can also contain reagents for preparing the reference population, such as, for example, adaptors, labels, polymerase, dNTP's, labelled dNTP's, PCR buffers, and the like, as well as printed instructions for preparing the reference population.

Flow Sorting of Microparticles with Up-Regulated and/or Down-Regulated Gene Products

After labeled polynucleotides are competitively hybridized to a reference population on microparticles, the microparticles may be analyzed and/or sorted in a number of ways depending on the chemical and/or physical properties of the microparticles and the attached sequences. For example, microparticles of interest may be mechanically separated by micro-manipulators, magnetic microparticles may be sorted by adjusting or manipulating magnetic fields, charged microparticles may be manipulated by electrophoresis, or the like. The following references provide

guidance for selecting means for analyzing and/or sorting microparticles: Pace, U.S. Patent 4,908,112; Saur et al., U.S. Patent 4,710,472; Senyei et al., U.S. Patent 4,230,685; Wilding et al., U.S. Patent 5,637,469; Penniman et al., U.S. Patent 4,661,225; Karnaukhov et al., U.S. Patent 4,354,114; Abbott et al., U.S. Patent 5,104,791; Gavin et al., PCT publication WO 97/40383; and the like. Preferably, microparticles containing fluorescently labeled DNA strands are conveniently classified and sorted by a commercially available FACS instrument, *e.g.* Van Dilla et al., Flow Cytometry: Instrumentation and Data Analysis (Academic Press, New York, 1985); Fulwyler et al., U.S. Patent 3,710,933; Gray et al., U.S. Patent 4,361,400; Dolbeare et al., U.S. Patent 4,812,394; and the like. For fluorescently labeled DNA strands competitively hybridized to a reference strand, preferably the FACS instrument has multiple fluorescent channel capabilities. Preferably, upon excitation with one or more high intensity light sources, such as a laser, a mercury arc lamp, or the like, each microparticle will generate fluorescent signals, usually fluorescence intensities, related to the quantity of labeled DNA strands from each cell or tissue types carried by the microparticle. As shown in Figure 1a of Example 1, when fluorescent intensities of each microparticle are plotted on a two-dimensional graph, microparticles indicating equal expression levels will be on or near the diagonal (100) of the graph. Up-regulated and down-regulated genes will appear in the off-diagonal regions (112). Such microparticles are readily sorted by commercial FACS instruments by graphically defining sorting parameters to enclose one or both off-diagonal regions (112) as shown in Figure 1b. Thus, microparticles can be sorted according to their relative optical signal, and if desired, collected for further analysis by accumulating those microparticles generating a signal within a predetermined range of values corresponding to a difference in gene expression among the different cell or tissue sources.

<u>Flow Sorting of Microparticles According to the Abundance of Nucleic Acid Sequences from which the Polynucleotides are Derived</u>

Microparticles containing fluorescently labeled DNA strands can also be classified and sorted according to the abundance of the gene products from which they are derived. The abundance of a nucleic acid sequence can be determined by the methods described above for determining relative gene expression and can be

correlated with the level of intensity of the optical signal generated by the polynucleotides bound to the microparticles. A lower intensity is indicative of a rarer nucleic acid sequence, such as a rare gene product. Rare genes are genes encoding an mRNA which is present in about 100 copies per cell or less, with increasing preference for less than about 50 copies to less than about 25 copies, with less than about 10 copies per cell being most preferred. Rare genes can be isolated by collecting microparticles with low fluorescent intensities as shown in Examples 9 and 10. The collected microparticles typically comprise less than about 5% of the total microparticles, with increasing preference for less than about 2.5%, 1%, to 0.5% with less than about 0.1% being most preferred.

Alternatively, since hybridization rates are proportionate to the abundance of a nucleic acid sequence, less abundant nucleic acid sequences can be isolated by setting the hybridization conditions such that nucleic acid sequences present in a lower abundance in a cell or tissue source remain unhybridized. Suitable hybridization conditions include those conditions used for producing normalized cDNA libraries (Patanjali et al., *Proc. Natl. Acad. Sci. USA*, 88:1943-1947 (1991)). For example, rare genes can be isolated by collecting unhybridized DNA after allowing a maximum period of time for hybridization of the abundant DNA species.

Repetitive sequences can often complicate the mapping and analysis of polymorphisms. Repetitive sequences exist due to the presence in the genome of transposons, retrotransposons, retroviruses, short interspersed repetitive elements (SINEs) such as Alu sequences, satellite DNA, minisatellite DNA, megasatellite DNA, and the like. Repetitive sequences can be removed from a DNA population as described above by sorting rapidly hybridizing DNA species away from DNA species that are slower to hybridize. Preferably, the unhybridized population is substantially enriched in polynucleotides derived from non-repetitive nucleic acid sequences.

Another aspect of the invention is a kit for analyzing and/or isolating nucleic acid sequences with respect to their abundance comprising microparticles prepared as described above and printed instructions for use.

Identification of Sorted Genes by Massively Parallel Signature Sequencing (MPSS)

Expressed genes may be identified in parallel by MPSS, which is a combination of two techniques: one for tagging and sorting fragments of DNA for parallel processing (*e.g.* Brenner et al., International application PCT/US96/09513), and another for the stepwise sequencing the end of a DNA fragment (*e.g.* Brenner, U.S. patent 5,599,675 and Albrecht et al., International patent application PCT/US97/09472). After an initial digestion of a target polynucleotide with a first restriction endonuclease, restriction fragments are ligated to oligonucleotide tags as described below, and in Brenner et al., International application PCT/US96/09513, so that the resulting tag-fragment conjugates may be sampled, amplified, and sorted onto separate solid phase supports by specific hybridization of the oligonucleotide tags with their tag complements.

Once an amplified sample of DNA fragments is sorted onto solid phase supports to form homogeneous populations of substantially identical fragments, the ends of the fragments are preferably sequenced with an adaptor-based method of DNA sequencing that includes repeated cycles of ligation, identification, and cleavage, such as the method described in Brenner, U.S. patent 5,599,675. In further preference, adaptors used in the sequencing method each have a protruding strand and an oligonucleotide tag selected from a minimally cross-hybridizing set of oligonucleotides, as taught by Albrecht et al., International patent application PCT/US97/09472. Such adaptors are referred to herein as "encoded adaptors." Encoded adaptors whose protruding strands form perfectly matched duplexes with the complementary protruding strands of a fragment are ligated. After ligation, the identity and ordering of the nucleotides in the protruding strand is determined, or "decoded," by specifically hybridizing a labeled tag complement, or "de-coder" to its corresponding tag on the ligated adaptor.

The preferred sequencing method is carried out with the following steps: (a) ligating an encoded adaptor to an end of a fragment, the encoded adaptor having a nuclease recognition site of a nuclease whose cleavage site is separate from its recognition site; (b) identifying one or more nucleotides at the end of the fragment by the identity of the encoded adaptor ligated thereto; (c) cleaving the fragment with a

nuclease recognizing the nuclease recognition site of the encoded adaptor such that the fragment is shortened by one or more nucleotides; and (d) repeating said steps (a) through (c) until said nucleotide sequence of the end of the fragment is determined. In the identification step, successive sets of tag complements, or "de-coders," are specifically hybridized to the respective tags carried by encoded adaptors ligated to the ends of the fragments. The type and sequence of nucleotides in the protruding strands of the polynucleotides are identified by the label carried by the specifically hybridized de-coder and the set from which the de-coder came, as described below.

Identification of Sorted Genes by Conventional Sequencing

Gene products carried by microparticles may be identified after sorting, *e.g.* by FACS, using conventional DNA sequencing protocols. Suitable templates for such sequencing may be generated in several different ways starting from the sorted microparticles carrying differentially expressed gene products. For example, the reference DNA attached to an isolated microparticle may be used to generate labeled extension products by cycle sequencing, *e.g.* as taught by Brenner, International application PCT/US95/12678. In this embodiment, primer binding site (400) is engineered into the reference DNA (402) distal to tag complement (406), as shown in Figure 4a. After isolating a microparticle, *e.g.* by sorting into separate microtiter well, or the like, the differentially expressed strands are melted off, primer (404) is added, and a conventional Sanger sequencing reaction is carried out so that labeled extension products are formed. These products are then separated by electrophoresis, or like techniques, for sequence determination. In a similar embodiment, sequencing templates may be produced without sorting individual microparticles. Primer binding sites (400) and (420) may be used to generate templates by PCR using primers (404) and (422). The resulting amplicons containing the templates are then cloned into a conventional sequencing vector, such as M13. After transfection, hosts are plated and individual clones are selected for sequencing.

In another embodiment, illustrated in Figure 4b, primer binding site (412) may be engineered into the competitively hybridized strands (410). This site need not have a complementary strand in the reference DNA (402). After sorting, competitively hybridized strands (410) are melted off of reference DNA (402) and amplified, *e.g.* by PCR, using primers (414) and (416), which may be labeled and/or derivatized with

biotin for easier manipulation. The melted and amplified strands are then cloned into a conventional sequencing vector, such as M13, which is used to transfect a host which, in turn, is plated. Individual colonies are picked for sequencing.

Example 1

Construction of a Tagged cDNA Library, Sampling, and Loading Tagged cDNAs onto Microparticles

In this example, a preferred protocol for preparing tagged reference DNA for loading onto microparticles is described. Briefly, cDNA from each of the cell or tissue types of interest is prepared and directionally cloned into a vector containing the tag element of Formula I. Preferably, the mRNA extracted from such cells or tissues is combined, usually in equal proportions, prior to first strand synthesis. mRNA is obtained using standard protocols, after which first and second strand synthesis is carried out as exemplified and the resulting cDNAs are inserted into a vector containing a tag element of Formula I, or like tag element. The vectors containing the tag-cDNA conjugates are then used to transform a suitable host, typically a conventional bacterial host, after which a sample of cells from the host culture is further expanded and vector DNA is extracted. The tag-cDNA conjugates are preferably amplified from the vectors by PCR and processed as described below for loading onto microparticles derivatized with tag complements. After the non-covalently attached strand is melted off, the cDNA-containing microparticles are ready to accept competitively hybridized gene products in accordance with the invention. Specific guidance relating to the indicated steps is available in Sambrook et al. (cited above); Ausbel et al., editors, Current Protocols in Molecular Biology (John Wiley & Sons, New York, 1995); and like guides on molecular biology techniques.

A pellet of approximately 5 μg of mRNA is resuspended in 45 μl (final volume) of a first strand pre-mix consisting of 10 μl 5x SuperScript buffer (250 mM Tris-Cl, pH 8.3, 375 mM KCl, and 15 mM $MgCl_2$) (GIBCO/BRL) (or like reverse transcriptase buffer), 5 μl 0.1 M dithiothreitol (DTT), 2.5 μl 3dNTP/methyl-dCTP mix (10 μM each of dATP, dGTP, dTTP, and 5-methyl-dCTP, e.g available from Pharmacia Biotech), 1 μl RNasin, 12 μl 0.25 μg/μl of reverse transcription primer shown below, and 14.5 μl H2O.

5'-biotin-GACATGCTGCATTGAGACGATTCTTTTTTTTTTTTTTTTTTV

Reverse Transcription Primer (SEQ ID NO: 2)

After incubation for 15 min at room temperature, 5 ml of 200 U/µl SuperScript is added and the mixture is incubated for 1 hr at 42°C. After the 1 hr incubation, the above mixture (about 50 µl total) is added to a second-strand premix on ice (volume 336 µl) consisting of 80 µl 5x second-strand buffer (94 mM Tris-Cl, pH 6.9, 453 mM KCl, 23 mM $MgCl_2$, and 50 mM $(NH_4)2SO_4$ to give a total reaction volume of about 386 µl. Separately, 4 µl of 0.8 U/µl RNase H (3.2 units) and 10 µl of 10 unit/µl E. coli DNA polymerase I (100 units) are combined and the combined enzyme mixture is added to the above second-strand reaction mixture, after which the total reaction volume is microfuged 5 sec and then incubated for 1 hr at 16°C and for 1 hr at room temperature to give the following double stranded cDNA (SEQ ID NO: 3):

```
5'-biotin-GACATGCTGCATTGAGACGATTCTTTTTTTTTTTTTTTTTTVXXX ... XGATCXXX-3'
          CTGTACGACGTAACTCTGCTAAGAAAAAAAAAAAAAAAAAABXXX ... XCTAGXXX-5'
                           ↑                                  ↑
                         Bsm BI                             Dpn II
```

where the X's indicated nucleotides in the cDNAs, V represents A, C, or G, and B represents C, G, or T. Note that the reverse transcription primer sequence has been selected to give a Bsm BI site in the cDNAs which results in a 5'-GCAT overhang upon digestion with Bsm BI.

After phenol/chloroform extraction and ethanol precipitation, the cDNA is resuspended in the manufacturer's recommended buffer for digestion with Dpn II (New England Biolabs, Beverely, MA), which is followed by capture of the biotinylated fragment on avidinated beads (Dynal, Oslo, Norway). After washing, the captured fragments are digested with Bsm BI to release the following cDNAs (SEQ ID NO: 4) which are precipitated in ethanol:

```
GCATTGAGACGATTCTTTTTTTTTTTTTTTTTTVXXX ... X    -3'
```

A<u>CTCTGC</u>TAAGAAAAAAAAAAAAAAAAAAABXXX ... XCTAG -5'

A conventional cloning vector, such as BlueScript II, pBC, or the like (Stratagene Cloning Systems, La Jolla, CA), is engineered to have the following sequence of elements (SEQ ID NO: 5)(which are those shown in Formula I):

```
5'-...TTAATTAAGGA [TAG] GGGCCCGCATAAGTCTTC [STUFFER]
GGATCC...-3'

3'-...AATTAATTCCT[TAG]CCCGGGCGTATTCAGAAG[STUFFER] CCTAGG...-5'
      ↑                                  ↑                 ↑
    Pac I                              Bbs I             Bam HI
```

After digestion with Bbs I and Bam HI, the vector is purified by gel electrophoresis and combined with the cDNAs for ligation. Note that the vector has been engineered so that the Bbs I digestion results in an end compatible with the Bsm BI-digested end of the cDNAs. After ligation, a suitable host bacteria is transformed and a culture is expanded for subsequent use.

From the expanded culture, a sample of host cells are plated to determine the fraction that carry vectors with inserted cDNAs, after which an aliquot of culture corresponding to about 1.7×10^5 insert-containing cells is withdrawn and separately expanded in culture. This represents about one percent of the repertoire of tags of the type illustrated in Formula I.

Preferably, the tag-cDNA conjugates are amplified out of the vectors by PCR using a conventional protocol, such as the following. For each of 8 replicate PCRs, the following reaction components are combined: 1 μl vector DNA (125 ng/μl for a library, 10^9 copies for a single clone); 10 μl 10x Klentaq Buffer (Clontech Laboratories, Palo Alto, CA); 0.25 μl biotinylated 20-mer "forward" PCR primer (1 nmol/μl); 0.25 μl FAM-labeled 20-mer "reverse" PCR primer (1 nmol/μl); 1 μl 25 mM dATP, dGTP, dTTP, and 5-methyl-dCTP (total dNTP concentration 100 mM); 5 μl DMSO; 2 μl 50x Klentaq enzyme; and 80.5 μl H_2O (for a total volume of 100 μl).

The PCR is run in an MJR DNA Engine (MJ Research), or like thermal cycler, with the following protocol: 1) 94°C for 4 min; 2) 94°C 30 sec; 3) 67°C 3 min; 4) 8 cycles of steps 2 and 3; 5) 94°C 30 sec, 6) 64°C 3 min, 7) 22 cycles of steps 5 and 6; 8) 67°C for 3 min; and 9) hold at 4°C.

The 8 PCR mixtures are pooled and 700 µl phenol is added at room temperature, after which the combined mixture is vortexed for 20-30 sec and then centrifuged at high speed (*e.g.* 14,000 rpm in an Eppendorf bench top centrifuge, or like instrument) for 3 min. The supernatant is removed and combined with 700 µl chloroform (24:1 mixture of chloroform:iso-amyl alcohol) in a new tube, vortexed for 20-30 sec, and centrifuged for 1 min, after which the supernatant is transferred to a new tube and combined with 80 µl 3M sodium acetate and 580 µl isopropanol. After centrifuging for 20 min, the supernatant is removed and 1 ml 70% ethanol is added. The mixture is centrifuged for 5-10 min, after which the ethanol is removed and the precipitated DNA is dried in a speedvac.

After resuspension, the cDNA is purified on avidinated magnetic beads (Dynal) using the manufacturer's recommended protocol and digested with Pac I (1 unit of enzyme per µg of DNA), also using the manufacturer's recommended protocol (New England Biolabs, Beverly, MA). The cleaved DNA is extracted with phenol/chloroform followed by ethanol precipitation. The tags of the tag-cDNA conjugates are rendered single stranded by combining 2 units of T4 DNA polymerase (New England Biolabs) per µg of streptavidin-purified DNA. 150 µg of streptavidin-purified DNA is resuspended in 200 µl H_2O and combined with the following reaction components: 30 µl 10 NEB Buffer No. 2 (New England Biolabs); 9 µl 100 mM dGTP; 30 µl T4 DNA polymerase (10 units/µl); and 31 µl H_2O; to give a final reaction volume of 300 µl. After incubation for 1 hr at 37°C, the reaction is stopped by adding 20 µl 0.5 M EDTA, and the T4 DNA polymerase is inactivated by incubating the reaction mixture for 20 min at 75°C. The tag-cDNA conjugates are purified by phenol/chloroform extraction and ethanol precipitation.

5 µm GMA beads with tag complements are prepared by combinatorial synthesis on an automated DNA synthesizer (Gene Assembler Special /4 Primers, Pharmacia Biotech, Bjorkgatan, Sweden, or like instrument) using conventional phosphoramidite chemistry, wherein nucleotides are condensed in the 3'→5' direction. In a preferred embodiment, a 28-nucleotide "spacer" sequence is synthesized,

followed by the tag complement sequence (8 "words" of 4 nucleotides each for a total of 32 nucleotides in the tag complement), and a sequence of three C's. Thus, the beads are derivatized with a 63-mer oligonucleotide. The length of the "spacer" sequence is not critical; however, the proximity of the bead surface may affect the activity of enzymes that are use to treat tag complements or captured sequences. Therefore, if such processing is employed, a spacer long enough to avoid such surface effects is desirable. Preferably, the spacer is between 10 and 30 nucleotides, inclusive. The following sequence (SEQ ID NO: 6), containing a Pac I site, is employed in the present embodiment:

```
5'-CCC-[Tag Complement]-TCCTTAATTAACTGGTCTCACTGTCGCA-bead
                                ↑
                              Pac I
```

Preferably, the tag-cDNA conjugates are hybridized to tag compliments on beads of a number corresponding to at least a full repertoire of tag complements, which in the case of the present embodiment is 8^8, or about 1.6×10^7 beads. The number of beads in a given volume is readily estimated with a hemocytometer.

Prior to hybridization of the tag-cDNA conjugates, the 5' ends of the tag complements are phosphorylated, preferably by treatment with a polynucleotide kinase. Briefly, 2.5×10^8 beads suspended in 100 μl H_2O are combined with 100 μl 10x NEB buffer No. 2 (New England Biolabs, Beverly, MA), 10 μl 100 mM ATP, 1 μl 10% Tween 20, 17 μl T4 polynucleotide kinase (10 units/μl), and 772 μl H_2O for a final volume of 1000 μl. After incubating for 2 hr at 37°C with vortexing, the temperature is increased to 65°C for 20 min to inactivate the kinase, with continued vortexing. After incubation, the beads are washed twice by spinning down the beads and resuspending them in 1 ml TE (Sambrook et al., Molecular Cloning, Second Edition, Cold Spring Harbor Laboratory) containing .01% Tween 20.

For hybridization of tag-cDNA conjugates to tag complements, the tag-cDNA conjugates as prepared above are suspended in 50 μl H_2O and the resulting mixture is combined with 40 μl 2.5x hybridization buffer, after which the combined mixture is filtered through a Spin-X spin column (0.22 μm) using a conventional protocol to give a filtrate containing the tag-cDNA conjugates. (5 ml of the 2.5x hybridization

buffer consists of 1.25 ml 0.1 M $NaPO_4$ (pH 7.2), 1.25 ml 5 M NaCl, 0.25 ml 0.5% Tween 20, 1.50 ml 25% dextran sulfate, and 0.75 ml H_2O.) Approximately 1.8 x 10^7 beads in 10 µl TE/Tween buffer (TE with .01% Tween 20) is centrifuged so that the beads form a pellet and the TE/Tween is removed. To the beads, 25 µl of 1x hybridization buffer (10 mM NaPO4 (pH 7.2), 500 mM NaCl, 0.01% Tween 20, 3% dextran sulfate) is added and the mixture is vortexed to fully resuspend the beads, after which the mixture is centrifuged so that the beads form a pellet and the supernatant is removed.

The tag-cDNA conjugates in the above filtrate are incubated at 75°C for 3 min and combined with the beads, after which the mixture is vortexed to fully resuspend the beads. The resulting mixture is further incubated at 75°C with vortexing for approximately three days (60 hours). After hybridization, the mixture is centrifuged for 2 min and the supernatant is removed, after which the beads are washed twice with 500 µl TE/Tween and resuspended in 500 µl 1x NEB buffer No. 2 with .01% Tween 20. The beads are incubated at 64°C in this solution for 30 min, after which the mixture is centrifuged so that the beads form a pellet, the supernatant is removed, and the beads are resuspended in 500 µl TE/Tween.

Loaded beads are sorted from unloaded beads using a high speed cell sorter, preferably a MoFlo flow cytometer equipped with an argon ion laser operating at 488 nm (Cytomation, Inc., Ft. Collins, CO), or like instrument. After sorting, the loaded beads are subjected to a fill-in reaction by combining them with the following reaction components: 10 µl 10x NEB buffer No. 2, 0.4 µl 25 mM dNTPs, 1 µl 1% Tween 20, 2 µl T4 DNA polymerase (10 units/ml), and 86.6 µl H_2O, for a final reaction volume of 100 µl. After incubation at 12°C for 30 min with vortexing, the reaction mixture is centrifuged so that the beads form a pellet and the supernatant is removed. The pelleted beads are resuspended in a ligation buffer consisting of 15 µl 10x NEB buffer No. 2, 1.5 µl 1% Tween 20, 1.5 µl 100 mM ATP, 1 µl T4 DNA ligase (400 units/ ml), and 131 µl H_2O, to give a final volume of 150 µl. The ligation reaction mixture is incubated at 37°C for 1 hr with vortexing, after which the beads are pelleted and washed once with 1x phosphate buffered saline (PBS) with 1 mM $CaCl_2$. The beads are resuspended in 45 µl PBS (with 1 mM $CaCl_2$) and combined with 6 µl Pronase solution (10 mg/ml, Boehringer Mannheim, Indianapolis, IN), after which the mixture is incubated at 37°C for 1 hr with vortexing. After centrifugation,

the loaded beads are washed twice with TE/Tween and then once with 1x NEB Dpn II buffer (New England Biolabs, Beverly, MA).

The tag-cDNA conjugates loaded onto beads are cleaved with Dpn II to produce a four-nucleotide protruding strand to which a complementary adaptor carrying a 3'-label is ligated. Accordingly, the loaded beads are added to a reaction mixture consisting of the following components: 10 μl 10x NEB Dpn II buffer, 1 μl 1% Tween, 4 μl Dpn II (50 units/ml), and 85 μl H_2O, to give a final reaction volume of 100 μl. The mixture is incubated at 37°C overnight with vortexing, after which the beads are pelleted, the supernatant is removed, and the beads are washed once with 1x NEB buffer No. 3. To prevent self-ligation, the protruding strands of the tag-cDNA conjugates are treated with a phosphatase, *e.g.* calf intestine phosphatase (CIP), to remove the 5' phosphates. Accordingly, the loaded beads are added to a reaction mixture consisting of the following components: 10 μl 10x NEB buffer No. 3, 1 μl 1% Tween 20, 5 μl CIP (10 units/μl), and 84 μl H_2O, to give a final reaction volume of 100 μl. The resulting mixture is incubated at 37°C for 1 hr with vortexing, after which the beads are pelleted, washed once in PBS containing 1 mM $CaCl_2$, treated with Pronase as described above, washed twice with TE/Tween, and once with 1x NEB buffer No. 2.

The following 3'-labeled adaptor (SEQ ID NO: 7) is prepared using conventional reagents, *e.g.* Clontech Laboratories (Palo Alto, CA):

```
5'-pGATCACGAGCTGCCAGTC-FAM
        TGCTCGACGGTCAG
```

where "p" is a 5' phosphate group and "FAM" is a fluorescein dye attached to the 3' carbon of the last nucleotide of the top strand by a commercially available 3' linker group (Clontech Laboratories). The ligation is carried out in the following reaction mixture: 5 μl 10x NEB buffer No. 2, 0.5 μl 1% Tween 20, 0.5 μl 100 mM ATP, 5 ml 3'-labeled adaptor (100 pmol/μl), 2.5 μl T4 DNA ligase (400 units/μl) and 36.5 μl H_2O, to give a final reaction volume of 50 μl. The reaction mixture is incubated at 16°C overnight with vortexing, after which the beads are washed once with PBS containing 1 mM $CaCl_2$ and treated with Pronase as described above. After this initial ligation, the nick remaining between the adaptor and tag-cDNA conjugate is sealed by

simultaneously treating with both a kinase and a ligase as follows. Loaded beads are resuspended in a reaction mixture consisting of the following components: 15 μ0x NEB buffer No. 2, 1.5 μl 1% Tween 20, 1.5 μl 100 mM ATP, 2 μl T4 polynucleotide kinase (10 units/μl), 1 μl T4 DNA ligase (400 units/μl), and 129 μl H_2O, for a final reaction volume of 150 μl.. The reaction mixture is incubated at 37°C for 1 hr with vortexing, after which the beads are washed once with PBS containing 1 mM $CaCl_2$, treated with Pronase as described above, and washed twice with TE/Tween.

After the labeled strand is melted off, preferably by treatment with 150 mM NaOH, the reference DNA on the beads is ready for competitive hybridization of differentially expressed gene products.

Example 2

Preparation of a Yeast Reference DNA Population
Attached to Microparticles

In this example, Saccharomyces cerevisiae cells of strain YJM920 MATa Gal+ SUC2 CUP1 are grown in separate rich and minimal media cultures essentially as describe by Wodicka et al. (cited above). mRNA extracted from cells grown under both conditions are used to establish a reference cDNA population which is tagged, sampled, amplified, labeled, and loaded onto microparticles. Loaded microparticles are isolated by FACS, labels are removed, and the non-covalently bound strands of the loaded DNA are melted off and removed.

Yeast cells are grown at 30°C either in rich medium consisting of YPD (yeast extract/peptone/glucose, Bufferad, Newark, NJ) or in minimal medium (yeast nitrogen base without amino acids, plus glucose, Bufferad). Cell density is measured by counting cells from duplicate dilutions, and the number of viable cells per milliliter is estimated by plating dilutions of the cultures on YPD agar immediately before collecting cells for mRNA extraction. Cells is mid-log phase (1-5 x 10^7 cells/ml) are pelleted, washed twice with AE buffer solution (50 mM NaAc, pH 5.2, 10 mM EDTA), frozen in a dry ice-ethanol bath, and stored at -80°C.

mRNA is extracted as follows for both the construction of the reference DNA library and for preparation of DNA for competitive hybridization. Total RNA is extracted from frozen cell pellets using a hot phenol method, described by Schmitt et al., Nucleic Acids Research, 18: 3091-3092 (1990), with the addition of a chloroform-

isoamyl alcohol extraction just befor precipitation of the total RNA. Phase-Lock Gel (5 Prime-3 Prime, Inc., Boulder, CO) is used for all organic extractions to increase RNA recovery and decrease the potential for contamination of the RNA with material from the organic interface. Poly(A)$^+$ RNA is purified from the total RNA with an oligo-dT selection step (Oligotex, Qiagen, Chatsworth, CA).

5 μg each of mRNA from cells grown on rich medium and minimal medium are mixed for construction of a cDNA library in a pUC19 containing the tag repertoire of Formula I. The tag repertoire of Formula I is digested with Eco RI and Bam HI and inserted into a similarly digested pUC19. The mRNA is reverse transcribed with a commercially available kit (Strategene, La Jolla, CA) using an olgio-dT primer containing a sequence which generates a Bsm BI site identical to that of Formula I upon second strand synthesis. The resulting cDNAs are cleaved with Bsm BI and Dpn II and inserted into the tag-containing pUC19 after digestion with Bsm BI and Bam HI. After transfection and colony formation, the density of pUC19 tranformants is determined so that a sample containing approximately thirty thousand tag-cDNA conjugates may be obtained and expanded in culture. Alternatively, a sample of tag-cDNA conjugates are obtained by picking approximately 30 thousand clones, which are then mixed and expanded in culture.

From a standard miniprep of plasmid, the tag-cDNA conjugates are amplified by PCR with 5-methyldeoxycytosine triphosphate substituted for deoxycytosine triphosphate. The following 19-mer forward and reverse primers (SEQ ID NO: 8 and SEQ ID NO: 9), specific for flanking sequences in pUC19, are used in the reaction:

```
forward primer:     5'-biotin-AGTGAATTCGGGCCTTAATTAA

reverse primer:     5'-FAM-GTACCCGCGGCCGCGGTCGACTCTAGAGGATC
```

where "FAM" is an NHS ester of fluorescein (Clontech Laboratories, Palo Alto, CA) coupled to the 5' end of the reverse primer via an amino linkage, *e.g.* Aminolinker II (Perkin-Elmer, Applied Biosystems Division, Foster City, CA). The reverse primer is selected so that a Not I site is reconstituted in the double stranded product. After PCR amplification, the tag-cDNA conjugates are isolated on avidinated beads, *e.g.* M-280 Dynabeads (Dynal, Oslo, Norway).

After washing, the cDNAs bound to the beads are digested with Pac I releasing the tag-cDNA conjugates and a stripping reaction is carried out to render the oligonucleotide tags single stranded. After the reaction is quenched, the tag-cDNA conjugate is purified by phenol-chloroform extraction and combined with 5.5 Om GMA beads carrying tag complements, each tag complement having a 5' phosphate. Hybridization is conducted under stringent conditions in the presence of a thermal stable ligase so that only tags forming perfectly matched duplexes with their complements are ligated. The GMA beads are washed and the loaded beads are concentrated by FACS sorting, using the fluorescently labeled cDNAs to identify loaded GMA beads. The isolated beads are treated with Pac I to remove the fluorescent label, after which the beads are heated in an NaOH solution using conventional protocols to remove the non-covalently bound strand. After several washes the GMA beads are ready for competitive hybridization.

Example 3

Isolation and Identification of Up-Regulated and Down-Regulated Genes in Yeast Exposed to Different Growth Conditions

In this example, mRNA is extracted from cells of each culture and two populations of labeled polynucleotides are produced by a single round of poly(dT) primer extension by a reverse transcriptase in the presence of fluorescently label nucleoside triphosphates. Equal amounts of each of the labeled polynucleotides are then combined with the GMA beads of Example 1 carrying the reference DNA population for competitive hybridization, after which the beads are analyzed by FACS and those in the off-diagonal regions are accumulated for MPSS analysis.

Fluorescent nucleoside triphosphates Cy3-dUTP or CY5-dUTP (Amersham) are incorporated into cDNAs during reverse transcription of 1. μg of poly(A)$^+$ RNA obtained as described in Example 1 using a poly(dT)$_{16}$ primer in separate reactions. After heating the primer and RNA to 70°C for 10 min, the reaction mixture is transferred to ice and a premixed solution, consisting of 200 U Superscript II (Gibco), buffer, deoxyribonucleoside triphosphates, and fluorescent nucleoside triphosphates are added to give the following concentrations: 500 μM for dATP, dCTP, and dGTP; 200 μM for dTTP; and 100 mM each for Cy3-dUTP or CY5-dUTP. After incubation at 42°C for 2 hours, unincorporated fluorescent nucleotides are removed by first

diluting the reaction mixture with 470 μl of 10 mM tris-HCl (pH 8.0)/1 mM EDTA. and then subsequently concentrating to about 5 μl using a Centricon-30 concentrator (Amicon). Purified labeled cDNA from both reactions is combined and resuspended in 11 μl of 3.5 x SSC containing 10 μg poly (dA) and 0.3 μl of 10% SDS. Prior to hybridization the solution is boiled for 2 min and allowed to cool to room temperature, after which it is applied to the GMA beads and incubated for about 8-12 hours at 62°C. After washing twice in 2 x SSC and 0.2% SDS, the GMA beads are resuspended in NEB-2 buffer (New England Biolabs, Beverly, MA) and loaded in a Coulter EPICS Elite ESP flow cytometer for analysis and sorting. In a two dimensional fluorescence intensity contour plot, the GMA beads generate a pattern as shown in Figure 1a. Sorting parameters are set as shown in Figure 1b so that GMA beads in the off-diagonal regions (112) are sorted and collected for MPSS analysis.

The labeled cDNA strands are melted from the GMA beads and removed by centrifugation. After several washes, a primer is annealed to the primer binding site shown in Formula I and extended in a conventional polymerization reaction to reconstitute the double stranded DNAs on the GMA beads which include the Dpn II site, described above. After digestion with Dpn II, beads loaded with tag-cDNA conjugates are placed in an instrument for MPSS analysis, as described in Albrecht et al. (cited above).

The top strands of the following 16 sets of 64 encoded adaptors (SEQ ID NO: 10 through SEQ ID NO: 25) are each separately synthesized on an automated DNA synthesizer (model 392 Applied Biosystems, Foster City) using standard methods. The bottom strand, which is the same for all adaptors, is synthesized separately then hybridized to the respective top strands:

SEQ ID NO.	Encoded Adaptor
10	5'-pANNNTACAGCTGCATCCCttggcgctgagg pATGCACGCGTAGGG-5'
11	5'-pNANNTACAGCTGCATCCCtgggcctgtaag pATGCACGCGTAGGG-5'
12	5'-pCNNNTACAGCTGCATCCCttgacgggtctc pATGCACGCGTAGGG-5'
13	5'-pNCNNTACAGCTGCATCCCtgcccgcacagt pATGCACGCGTAGGG-5'
14	5'-pGNNNTACAGCTGCATCCCttcgcctcggac

```
                pATGCACGCGTAGGG-5'
15      5'-pNGNNTACAGCTGCATCCCtgatccgctagc
                pATGCACGCGTAGGG-5'
16      5'-pTNNNTACAGCTGCATCCCttccgaacccgc
                pATGCACGCGTAGGG-5'
17      5'-pNTNNTACAGCTGCATCCCtgagggggatag
                pATGCACGCGTAGGG-5'
18      5'-pNNANTACAGCTGCATCCCttcccgctacac
                pATGCACGCGTAGGG-5'
19      5'-pNNNATACAGCTGCATCCCtgactccccgag
                pATGCACGCGTAGGG-5'
20      5'-pNNCNTACAGCTGCATCCCtgtgttgcgcgg
                pATGCACGCGTAGGG-5'
21      5'-pNNNCTACAGCTGCATCCCtctacagcagcg
                pATGCACGCGTAGGG-5'
22      5'-pNNGNTACAGCTGCATCCCtgtcgcgtcgtt
                pATGCACGCGTAGGG-5'
23      5'-pNNNGTACAGCTGCATCCCtcggagcaacct
                pATGCACGCGTAGGG-5'
24      5'-pNNTNTACAGCTGCATCCCtggtgaccgtag
                pATGCACGCGTAGGG-5'
25      5'-pNNNTTACAGCTGCATCCCtcccctgtcgga
                pATGCACGCGTAGGG-5'
```

where N is any of dA, dC, dG, or dT; p is a phosphate group; and the nucleotides indicated in lower case letters are the 12-mer oligonucleotide tags. Each tag differs from every other by 6 nucleotides. Equal molar quantities of each adaptor are combined in NEB #2 restriction buffer (New England Biolabs, Beverly, MA) to form a mixture at a concentration of 1000 pmol/μL.

Each of the 16 tag complements are separately synthesized as amino-derivatized oligonucleotides and are each labeled with a fluorescein molecule (using an NHS-ester of fluorescein, available from Molecular Probes, Eugene, OR) which is attached to the 5' end of the tag complement through a polyethylene glycol linker (Clonetech Laboratories, Palo Alto, CA). The sequences of the tag complements are simply the 12-mer complements of the tags listed above.

Ligation of the adaptors to the target polynucleotide is carried out in a mixture consisting of 5 μl beads (20 mg), 3 μL NEB 10x ligase buffer, 5 μL adaptor mix (25 nM), 2.5 μL NEB T4 DNA ligase (2000 units/μL), and 14.5 μL distilled water. The

mixture is incubated at 16°C for 30 minutes, after which the beads are washed 3 times in TE (pH 8.0).

After centrifugation and removal of TE, the 3' phosphates of the ligated adaptors are removed by treating the polynucleotide-bead mixture with calf intestinal alkaline phosphatase (CIP) (New England Biolabs, Beverly, MA), using the manufacturer's protocol. After removal of the 3' phosphates, the CIP may be inactivated by proteolytic digestion, *e.g.* using PronaseTM (available form Boeringer Mannhiem, Indianapolis, IN), or an equivalent protease, with the manufacturer's protocol. The polynucleotide-bead mixture is then washed, treated with a mixture of T4 polynucleotide kinase and T4 DNA ligase (New England Biolabs, Beverly, MA) to add a 5' phosphate at the gap between the target polynucleotide and the adaptor, and to complete the ligation of the adaptors to the target polynucleotide. The bead-polynucleotide mixture is then washed in TE.

Separately, each of the labeled tag complements is applied to the polynucleotide-bead mixture under conditions which permit the formation of perfectly matched duplexes only between the oligonucleotide tags and their respective complements, after which the mixture is washed under stringent conditions, and the presence or absence of a fluorescent signal is measured. Tag complements are applied in a solution consisting of 25 nM tag complement 50 mM NaCl, 3 mM Mg, 10 mM Tris-HCl (pH 8.5), at 20°C, incubated for 10 minutes, then washed in the same solution (without tag complement) for 10 minute at 55°C.

After the four nucleotides are identified as described above, the encoded adaptors are cleaved from the polynucleotides with Bbv I using the manufacturer's protocol. After an initial ligation and identification, the cycle of ligation, identification, and cleavage is repeated three times to give the sequence of the 16 terminal nucleotides of the target polynucleotide.

Preferably, analysis of the hybridized encoded adaptors takes place in an instrument which i) constrains the loaded microparticles to be disposed in a planar array in a flow chamber, ii) permits the programmed delivery of process reagents to the flow chamber, and iii) detects simultaneously optical signals from the array of microparticles. Such a preferred instrument is shown diagrammatically in Figure 2, and more fully disclosed in Bridgham et al., International patent application PCT/US98/11224. Briefly, flow chamber (500) is prepared by etching a cavity

having a fluid inlet (502) and outlet (504) in a glass plate (506) using standard micromachining techniques, *e.g.* Ekstrom et al., International patent application PCT/SE91/00327; Brown, U.S. patent 4,911,782; Harrison et al., Anal. Chem. 64: 1926-1932 (1992); and the like. The dimension of flow chamber (500) are such that loaded microparticles (508), *e.g.* GMA beads, may be disposed in cavity (510) in a closely packed planar monolayer of 100-200 thousand beads. Cavity (510) is made into a closed chamber with inlet and outlet by anodic bonding of a glass cover slip (512) onto the etched glass plate (506), *e.g.* Pomerantz, U.S. patent 3,397,279. Reagents are metered into the flow chamber from syringe pumps (514 through 520) through valve block (522) controlled by a microprocessor as is commonly used on automated DNA and peptide synthesizers, *e.g.* Bridgham et al., U.S. patent 4,668,479; Hood et al., U.S. patent 4,252,769; Barstow et al., U.S. patent 5,203,368; Hunkapiller, U.S. patent 4,703,913; or the like.

Three cycles of ligation, identification, and cleavage are carried out in flow chamber (500) to give the sequences of 12 nucleotides at the termini of each of approximately 100,000 fragments. Nucleotides of the fragments are identified by hybridizing tag complements to the encoded adaptors as described above. Specifically hybridized tag complements are detected by exciting their fluorescent labels with illumination beam (524) from light source (526), which may be a laser, mercury arc lamp, or the like. Illumination beam (524) passes through filter (528) and excites the fluorescent labels on tag complements specifically hybridized to encoded adaptors in flow chamber (500). Resulting fluorescence (530) is collected by confocal microscope (532), passed through filter (534), and directed to CCD camera (536), which creates an electronic image of the bead array for processing and analysis by workstation (538). Preferably, after each ligation and cleavage step, the cDNAs are treated with PronaseTM or like enzyme. Encoded adaptors and T4 DNA ligase (Promega, Madison, WI) at about 0.75 units per μL are passed through the flow chamber at a flow rate of about 1-2 μL per minute for about 20-30 minutes at 16°C, after which 3' phosphates are removed from the adaptors and the cDNAs prepared for second strand ligation by passing a mixture of alkaline phosphatase (New England Bioscience, Beverly, MA) at 0.02 units per μL and T4 DNA kinase (New England Bioscience, Beverly, MA) at 7 units per μL through the flow chamber at 37°C with a flow rate of 1-2 μL per minute for 15-20 minutes. Ligation is accomplished by T4

DNA ligase (.75 units per mL, Promega) through the flow chamber for 20-30 minutes. Tag complements at 25 nM concentration are passed through the flow chamber at a flow rate of 1-2 μL per minute for 10 minutes at 20°C, after which fluorescent labels carried by the tag complements are illuminated and fluorescence is collected. The tag complements are melted from the encoded adaptors by passing hybridization buffer through the flow chamber at a flow rate of 1-2 μL per minute at 55°C for 10 minutes. Encoded adaptors are cleaved from the cDNAs by passing Bbv I (New England Biosciences, Beverly, MA) at 1 unit/μL at a flow rate of 1-2 μL per minute for 20 minutes at 37°C.

Example 4

FACS Analysis of Microparticles Loaded with Different Ratios of DNAs Labeled with Fluorescein and CY5

In this example, the sensitivity of detecting different ratios of differently labeled cDNAs was tested by constructing a reference DNA population consisting of a single clone and then competitively hybridizing to the reference DNA population different ratios of complementary strands labeled with different fluorescent dyes. The reference DNA population consisted of a cDNA clone, designated "88.11," which is an 87-basepair fragment of an expressed gene of the human monocyte cell line THP-1, available from the American Type Culture Collection (Rockville, Maryland) under accession number TIB 202. The nucleotide sequence of 88.11 has a high degree of homology to many entries in the GenBank Expressed Sequence Tag library, *e.g.* gb AA830602 (98%). The reference DNA population, which consisted of only 88.11 cDNA, was prepared as described in Example 1, with the exception that a special population of microparticles was prepared in which all microparticles had the same tag complement attached. The corresponding oligonucleotide tag was attached to the 88.11 cDNA. Thus, only monospecific populations of tags and tag complements were involved in the experiment. After competitive hybridization, the loaded microparticles were analyzed on a Cytomation, Inc. (Ft. Collins, CO) FACS instrument as described above.

88.11 cDNA was also cloned into a vector identical to that of Example 1 (330 of Figure 3b), except that it did not contain tag 336. 10 μg of vector DNA was linearized by cleaving to completion with Sau 3A, an isoschizomer of Dpn II (342 of

Figure 3b), after which two 1 μg aliquots of the purified linear DNA were taken. From each 1 μg aliquot, about 20 μg of labeled single stranded DNA product was produced by repeated cycles of linear amplification using primers specific for primer binding site 332. In one aliquot, product was labeled by incorporation of rhodamine R110-labeled dUTP (PE Applied Biosystems, Foster City, CA); and in the other aliquot, product was labeled by incorporation of CY5-labeled dUTP (Amersham Corporation, Arlington Heights, IL). Quantities of the labeled products were combined to form seven 5 μg amounts of the two products in ratios of 1:1, 2:1, 1:2, 4:1, 1:4, 8:1, and 1:8. The 5 μg quantities of labeled product were separately hybridized to 1.6 x 10^5 microparticles (GMA beads with 88.11 cDNA attached) overnight at 65°C in 50 μl 4x SSC with 0.2% SDS, after which the reaction was quenched by diluting to 10 ml with ice-cold TE/Tween buffer (defined above). The loaded microparticles were centrifuged, washed by suspending in 0.5 ml 1x SSC with 0.2% SDS for 15 min at 65°C, centrifuged, and washed again by suspending in 0.5 ml 0.1x SSC with 0.2% SDS for 15 min at 55°C. After the second washing, the microparticles were centrifuged and resuspended in 0.5 ml TE/Tween solution for FACS analysis.

The results are shown in Figures 5a-5e, where in each Figure the vertical axis corresponds to CY5 fluorescence and the horizontal axis corresponds to rhodamine R110 fluorescence. In Figure 5a, a population of microparticles were combined that had either all R110-labeled DNA or all CY5-labeled DNA hybridized to the complementary reference strands. Contours 550 and 552 are clearly distinguished by the detection system of the FACS instrument and microparticles of both populations produce readily detectable signals. Figure 5b illustrates the case where the R110- and CY5-labeled strands are hybridized in equal proportions. As expected, the resulting contour is located on the diagonal of the graph and corresponds to the position expected for non-regulated genes. Figures 5c through 5e show the analysis of three pairs of competitive hybridizations: i) R110- and CY5-labeled strands hybridized in a 2:1 concentration ratio and a 1:2 concentration ratio, ii) R110- and CY5-labeled strands hybridized in a 4:1 concentration ratio and a 1:4 concentration ratio, and iii) R110- and CY5-labeled strands hybridized in an 8:1 concentration ratio and a 1:8 concentration ratio. The data of Figure 5c suggest that genes up-regulated or down-regulated by a factor of two are detectable in the present embodiment, but that

significant overlap may exist between signals generated by regulated and non-regulated genes. Figures 5d and 5e suggest that genes up-regulated or down-regulated by a factor of four or higher are readily detectable over non-regulated genes.

Example 5

FACS Analysis of Differentially Expressed Genes from Stimulated and Unstimulated THP-1 Cells

In this example, a reference DNA population attached to microparticles was constructed from cDNA derived from THP-1 cells stimulated as indicated below. Equal concentrations of labeled cDNAs from both stimulated and unstimulated THP-1 cells were then competitively hybridized to the reference DNA population, as described in Example 1, and the microparticles carrying the labeled cDNAs were analyzed by a FACS instrument. THP-1 cells were stimulated by treatment with phorbol 12-myristate 13-acetate (PMA) and lipopolysaccharide (LPS).

THP-1 cells were grown in T-165 flasks (Costar, No. 3151) containing 50 ml DMEM/F12 media (Gibco, No. 11320-033) supplemented with 10% fetal bovine serum (FBS)(Gibco, No. 26140-038), 100 units/ml penicillin, 100 μg/ml streptomycin (Gibco, No. 15140-122), and 0.5 μM β-mercaptoethanol (Sigma, No. M3148). Cultures were seeded with 1×10^5 cells/ml and grown to a maximal density of 1×10^6. Doubling time of the cell populations in culture was about 36 hours. Cells were treated with PMA as follows: Cells from a flask (about 5×10^7 cells) were centrifuged (Beckman model GS-6R) at 1200 rpm for 5 minutes and resuspended in 50 ml of fresh culture media (without antibiotics) containing 5 μl of 1.0 mM PMA (Sigma, No. P-8139) in DMSO (Gibco No. 21985-023) or 5 μl DMSO (for the unstimulated population), after which the cells were cultured for 48 hours. Following the 48 hour incubation, media and non-adherent cells were aspirated from the experimental flask (*i.e.* containing stimulated cells) and fresh media (without antibiotics) was added, the fresh media containing 10 μl of 5 mg/ml LPS (Sigma, No. L-4130) in phosphate buffered saline (PBS). The culture of unstimulated cells was centrifuged (Beckman model GS-6R) at 1200 rpm for 5 minutes at 4°C so that a pellet formed which was then resuspended in 50 ml of fresh growth media containing 10 μl PBS. Both the cultures of stimulated and unstimulated cells were incubated at 37°C for four hours, after which cells were harvested as follows: Media was aspirated from

the cultures and adherent cells were washed twice with warm PBS, after which 10 ml PBS was added and the cells were dislodged with a cell scaper. The dislodged cells were collected and their concentration was determined with a hemocytometer, after which they were centrifuged (Beckman model GS-6R) at 1200 rpm for 5 minutes to form a pellet which was used immediately for RNA extraction.

mRNA was extracted from about 5 x 10^6 cells using a FastTrack 2.0 kit (No. K1593-02, Invitrogen, Inc. San Diego, CA) for isolating mRNA. The manufacturer's protocol was followed without significant alterations. A reference DNA population attached to microparticles was constructed from mRNA extracted from stimulated cells, as described in Example 1. Separate cDNA libraries were constructed from mRNA extracted from stimulated and unstimulated cells. The vectors used for the libraries were identical to that of Example 1, except that they did not contain oligonucleotide tags (336 of Figure 3b). Following the protocol of Example 4, approximately 2.5 μg of rhodamine R110-labeled single stranded DNA was produced from the cDNA library derived from stimulated cells, and approximately 2.5 μg of CY5-labeled single stranded DNA was produced from the cDNA library derived from unstimulated cells. The two 2.5 μg aliquots were mixed and competitively hybridized to the reference DNA on 9.34 x 10^5 microparticles. The reaction conditions and protocol was as described in Example 4.

After hybridization, the microparticles were sorted by a Cytomation, Inc. MoFlo FACS instrument as described above. Figure 6 contains a conventional FACS contour plot 600 of the frequencies of microparticles with different fluorescent intensity values for the two fluorescent dyes. Approximately 10,000 microparticles corresponding to up-regulated genes (sort window 602 of Figure 6) were isolated, and approximately 12,000 microparticles corresponding to down-regulated genes (sort window 604 of Figure 6) were isolated. After melting off the labeled strands, as described above, the cDNAs carried by the microparticles were amplified using a commercial PCR cloning kit (Clontech Laboratories, Palo Alto, CA), and cloned into the manufacturer's recommended cloning vector. After transformation, expansion of a host culture, and plating, 87 colonies of up-regulated cDNAs were picked and 73 colonies of down-regulated cDNAs were picked. cDNAs carried by plasmids extracted from these colonies were sequenced using conventional protocols on a PE

Applied Biosystems model 373 automated DNA sequencer. The identified sequences are listed in Tables 1 and 2.

Table 1

Up-Regulated Genes

No. Copies	Description	GenBank Indentifier
19	LD78/MIP-1	HUMCKLD78
16	TNF-inducible (TSF-6) mRNA	HUMTSG6A
15	GRO-γ (MIP-2β)	HUMGROG5
6	GRO-β (MIP-2α)	HUMGROB
6	act-2	HUMACT2A
4	guanylate binding protein isoform I (GBP-2)	HUMGBP1
4	spermidine/spermin N1-acetyltransferase	HUMSPERMNA
4	adipocyte lipid-binding protein	HUMALBP
3	Fibronectin	HSFIB1
3	interleukin-8	HSMDNCF
1	insulin-like growth factor binding protein 3	HSIGFBP3M
1	interferon-γ inducible early response gene	HSINFGER
1	type IV collangenase	
1	cathepsin L	HSCATHL
1	EST	
1	EST	
1	Genomic/EST	HSAC002079

Table 2

Down-Regulated Genes

No. Copies	Description	GenBank Indentifier
16	Elongation factor 1	HSEF1AC
4	Ribosomal protein S3a/v-fos tranf. Effector	HUMFTE1A
6	Ribosomal protein S7	HUMRPS17
2	Translationally controlled tumor protein	HSTUMP
3	23 kD highly basic protein	HS23KDHBP
2	Laminin receptor	HUMLAMR
2	Cytoskeletal gamma-actin	HSACTCGR
2	Ribosomal protein L6	HSRPL6AA
2	Ribosomal protein L10	HUMRP10A
2	Ribosomal protein L21	HSU14967
2	Ribosomal protein S27	HSU57847
1	Ribosomal protein L5	HSU14966
1	Ribosomal protein L9	HSU09953
1	Ribosomal protein L17	HSRPL17
1	Ribosomal protein L30	HSRPL30
1	Ribosomal protein L38	HSRPL38
1	Ribosomal protein S8	HSRPS8
1	Ribosomal protein S13	HSRPS13
1	Ribosomal protein S18	HSRPS18
1	Ribosomal protein S20	HUMRPS20
1	Acidic ribosomal phosphoprotein P0	HUMPPARP0
1	26S proteasome subunit p97	HUM26SPSP
1	DNA-binding protein B	HUMAAE
1	T-cell cyclophilin	HSCYCR
1	Interferon inducible 6-26 mRNA	HSIFNIN4
1	Hematopoetic proteoglycan core protein	HSHPCP
1	Fau	HSFAU
1	beta-actin	HSACTB
1	Nuclear enc. mito. serine hydroxymethyltrans.	HUMSHMTB
1	Mito. Cytochrome c oxidase subunit II	HUMMTCDK
1	Genomic	W92931
1	EST	W84529
1	EST	AA933890
1	EST	AA206288
1	EST	AA649735
1	EST	N34678
1	EST	AA166702
1	EST	AA630799
3	Genomic	

Example 6

FACS Analysis of Differentially Expressed Genes from Stimulated and Unstimulated THP-1 Cells

(Experiment: Comp 11)

A reference DNA population attached to microparticles was constructed from cDNA derived from stimulated THP-1 cells. cDNA from stimulated and unstimulated THP-1 cells was prepared for competitive hybridization as follows. 20 μg each of the THP-1 unstimulated probe library (U3A-TL) and the THP-1 stimulated probe library (S3A-TL) were digested with 50 units of Sau3A to prepare the vector for linear PCR. The DNA was purified by phenol/chloroform extraction and fluorescently labelled by PCR. For calibration purposes, both CY5 and R110 were used to label each condition.

The U3A-TL DNA was labeled with CY5 and the S3A-TL DNA was labeled with R110. Briefly, a reaction mixture containing 80 μl 10X PCR Buffer; 16 μl biotinylated primer (B-Primer, 125 pmole/:l); 16 μl dNTPs (6.25 mM); 4 μg template; 16 μl Klentaq enzyme; 64 μl R110 dUTP or 6.4 μl of CY5 dUTP; and water to bring the total volume to 800 μl. This mixture was dispensed into 8 aliquots, which then underwent 34 cycles of PCR according to the following protocol: 1) 94°C 3 min; 2) 94°C 30 sec; 3) 62°C 30 sec; 4) 72°C 1 min; and 5) 72°C 10 min. The PCR reaction was purified and the colored nucleotides were removed by precipitation.

Reference Population

The Comp 11 bead library consisted of 2,667,369 beads, with a complexity of 1 million clones from the THP-1 stimulated library. The beads were prepared as described above as outlined in Figure 3. The starting PMT2 mean for the FITC signal was 19.5. The duplexed DNA on the beads was denatured with 2.5 ml 150mM NaOH washes at RT for 15min with mild vortexing. The efficiency of the denaturization was determined by measuring the remaining FITC signal mean, which was 2.2, i.e., 11.3% residual fluorescence. The beads were washed twice in .5 ml of 4X SSC .1% SDS.

Competitive Hybridization

100,000 beads were hybridized with 10 μg of each linear PCR product of the stimulated probe library (S3A-TL) labeled with CY5 and the same library labeled with R110. 936,542 beads were hybridized with 10 μg of CY5 stimulated probe and 10 μg of R110 unstimulated probe. The beads were assembled in 50 μl with a final

buffer composition of 4X SSC/.1% SDS. The samples were heated to 80°C for 3 minutes, the probes were added and the temperature was moved to 65°C. Hybridization continued for 16 hrs. with vortexing. The beads were ice quenched in 10 ml of TE Tween. The recovered samples were rinsed 2 times with 1X SSC /.1% SDS, resuspended in .5 ml of 1X SSC /.1%SDS, and washed at 65°C for 15 min. The beads were rinsed in .1X SSC /.1%SDS and washed at 55°C in .1X SSC /.1% SDS for 15 min. The samples were rinsed with TE Tween and 10,000 events of both samples were analyzed on the BD FacsCaliber. 10,163 beads (1.15%), the brightest CY5 off the 1:1 diagonal, were sorted. 11,977 beads (1.35%), the brightest R110 off the 1:1 diagonal, were sorted. The beads were pooled in a PCR reaction, TA cloned, and sequenced. The identified sequences are listed in Tables 3 and 4.

Table 3

Comp 11: Downregulated Genes

No. Copies	Description	GenBank Identifier
99	23 kD highly basic protein	HS23KDHBP
1	26S proteasome subunit p55	AB003103
7	26S proteasome subunit p97	HUM26SPSP
1	28kD heat shock protein	HSHSP28
3	90 kD HSP	HSHSP90R
1	α NAC	HSANAC
2	α-enolase	HSAEP
3	α1 acid glycoprotein	HUMAGP1A
21	Acidic ribosomal phosphoprotein P0	HUMPPARP0
4	Acidic ribosomal phosphoprotein P1	HUMPPARP1
3	Acidic ribosomal phosphoprotein P2	HUMPPARP2
1	activin β-C chain	HSACTNBC
3	Adenylyl cyclase-associated protein (CAP)	HUMADCY
2	ADP/ATP translocase	HUMTLCA
3	Allograft-inflammatory factor 1	HSU19713
13	Antioxidant enzyme AOE37-2	HSU25182
1	Arp2/3 protein complex subunit p21Arc	AF006084
2	Arp2/3 protein complex subunit p41Arc	AF06086
1	ATP-dependent RNA helicase	AB001636
1	B94	HUMB94
7	basic transcription factor 3a (BTF3a)	HSBTF3
3	BBC1	HSBBC1
3	beta-actin	HSACTB
1	brain-expressed HHCPA78 homolog	S73591
1	c-myc transcription factor puf	HUMPUF
3	Calmodulin	HUMCAMA
1	cAMP response element regulatory protein	HUMCREB2A

No. Copies	Description	GenBank Identifier
1	cis-acting sequence	HUMCIS
1	Cks1 protein homolog	HSCKSNS2
3	clathrin assembly protein 50	HSU36188
1	CLP	HUMCLPB
5	Cu/Zn SOD-1	HSSODR1
1	Cyclophilin	HUMCYCLO
3	Cytochrome c oxidase cox VIIa-L	HSCOX7AL
1	Cytochrome c oxidase subunit Vb	HUMCOXCA
3	Cytochrome c oxidase subunit Vic	HSCOVIC
4	Cytoskeletal gamma-actin	HSACTCGR
3	Cytoskeletal tropomycin TM30	HSTROPCR
4	DNA-binding protein B	HUMAAE
2	EBV small RNAs associated protein (EAP)	HSEAP
30	Elongation factor 1 α	HSEF1AC
1	Elongation factor 1 δ	HSEF1DELA
1	Elongation factor 1 γ	HSEF1GMR
1	ERp28 protein	HSERP28
9	Fau	HSFAU
1	ferritin L chain	HUMFERL
1	Fibronectin receptor	HSFNRA
4	Fus	HSFUSA
2	G-β-like protein	HUMMHBA123
4	Glutaminyl tRNA synthetase	HSGTS
2	H+ ATP synthase subunit b	HSATPSYN
2	H3.3 histone, class B	HUMHISH3B
10	Heat shock factor binding protein 1	AF068754
5	Heat shock protein 86	HSHSP86
4	Hematopoetic lineage cell specific protein	HSHEAM
2	Hematopoetic proteoglycan core protein	HSHPCP
1	HLA-DR associated protein	HSPHAPII
3	HMG-17	HUMHMG17
2	Icln chlorine channel regulatory protein	HSU17899
2	IL-8	HSMDNCF
1	IMP dehydrogenase	HUMIMP
3	Initiation factor 4B	HSINTFA4B
1	Insulinoma rig analog	HUMIDB
2	Interferon inducible 6-26 mRNA	HSIFNIN4
1	KIAA0116	HUMORFA10
1	KIAA0164	D79986
5	KIAA0571	AB01114B
11	Lactate dehydrogenase B	HSLDHBR
12	Laminin receptor	HUMLAMR
7	LD78/MIP-1	HUMCKLD78
1	Leucine-rich protein	HUM130LEU
1	LLRep3	HSLLREP3
1	low Mr GTP-binding protein (RAB32)	HSU71127

No. Copies	Description	GenBank Identifier
1	LST1	HSLST1G
2	MAPKAP kinase (3pK)	HSU09578
5	MHC protein hom. to chicken B complex protein	HUMMHBA123
1	Mitochondrial cytochrome c oxidase subunit II	HUMMTCDK
2	Mitochondrial phosphate carrier protein	SSMPCP
1	Mitochondrial serine hydroxymethyl transferase	HUMSHMTB
1	Mitochondrial tRNAs	MIT1HS
2	Mitochondrial ubiquinone-binding protein	HSUBPQPC
1	Mn SOD-2	HUMSUDIS
2	Myeloid progenitor inhibitory factor (MPIF-1)	HSU85767
1	Myosin regulatory light chain	HSMRLCM
1	Nuclear-encoded mito. serine hydroxymethyltransferase	HUMSHMTB
1	P2U nucleotide receptor	S74902
8	Palmitoyl-protein thioesterase	HSU44772
1	Phosphate carrier	HSPHOSC
2	Prothymosin α	HUMTHYMA
21	Ribosomal protein L10	HUMRP10A
4	ribosomal protein L11	HSRPL11
8	ribosomal protein L14	D87735
13	ribosomal protein L17	HSRPL17
4	ribosomal protein L18a	HUMRIBPROD
27	ribosomal protein L21	HSU14967
4	ribosomal protein L23 (putative)	HSL23MR
4	ribosomal protein L25	HSU12465
4	ribosomal protein L26	HSRP26AA
6	ribosomal protein L27a	HSU14968
1	ribosomal protein L28	HSU14969
15	ribosomal protein L29	HSRPL29
10	ribosomal protein L3	HUMRRL3A
2	ribosomal protein L30	HUMRPL30A
1	ribosomal protein L30	HSRPL30
12	ribosomal protein L32	HSRPL32
1	ribosomal protein L34	HUMRPL34A
1	ribosomal protein L35	HSU12465
3	ribosomal protein L37a	HSRPL37A
11	ribosomal protein L38	HSRPL38
8	ribosomal protein L4	HSHRPL4
3	ribosomal protein L41	AF026844
19	ribosomal protein L5	HSU14966
22	ribosomal protein L6	HSRPL6AA
17	ribosomal protein L7	HSRBPRL7A
13	ribosomal protein L7a	HUMRPL7A
19	ribosomal protein L9	HSU09953

No. Copies	Description	GenBank Identifier
25	ribosomal protein S11	HSRPS11
5	ribosomal protein S13	HSRPS13
11	ribosomal protein S15a	HSRPS15A
5	ribosomal protein S16	HUMSRAA
28	ribosomal protein S17	HUMRPS17
35	ribosomal protein S18	HSRPS18
2	ribosomal protein S19	HUMS19RP
6	ribosomal protein S20	HUMRPS20
11	ribosomal protein S27	HSU57847
17	ribosomal protein S28(hu homolog of yeast)	HUMRSPT
3	ribosomal protein S3	HSHUMS3
4	ribosomal protein S3a/v-fos transf. effector protein	HUMFTE1A
6	ribosomal protein S4	HUMRPS4X
1	ribosomal protein S7	HUMRPS7A
20	ribosomal protein S8	HSRPS8
1	RNAse/angiogenin inhibitor	HSRAI
1	small nuclear RNA U2	HSU25766
5	T-cell cyclophilin	HSCYCR
1	T-cell surface glycoprotein	HSE2
1	TI-227H	HUMTI227HC
1	transcriptional coactivator PC4	HSU12979
1	translation initiation factor 2 β subunit	HUMELF2
1	translation initiation factor eIF3 p40 subunit	HSU54559
2	translationally controlled tumor protein	HSTUMP
1	U1 small nuclear RNP-specific C protein	HSU1RNPC
2	ubiquinol-cytochrome c oxidase smallest subunit	D55636
1	ubiquinone binding protein	HUMQBPCA
4	Ubiquitin	HSU49869
2	Ubiquitin	HUMUBI13
7	ubiquitin Uba52	HSUBA52P
11	ubiquitin Uba80	HSIBA80R
8	EST	AA149853
6	EST	AA759306
3	EST	AI053510
2	EST	AA630799
2	EST	N26031
2	EST	AA843411
2	EST	AA234913
2	EST	AI034446
2	EST	AI054090
1	EST	AI054090
1	EST	AA828574
1	EST	AI087086
1	EST	AI031866

No. Copies	Description	GenBank Identifier
1	EST	AI040041
1	EST	AA542832
1	EST	N73319
1	EST	AA464447
1	EST	AA993034
1	EST	AA933890
1	EST	AI095923
1	EST	AA166703
1	EST	AA573139
1	EST	AA938293
1	EST	H43222
1	EST	AA576961
1	EST	AI015173
1	EST	AI015700
1	EST	W95680
1	EST	AA934688
1	EST	AA204731
2	NO MATCH	
1	NO MATCH	
1	NO MATCH	
1	NO MATCH	
918	**total sequenced (downregulated)**	

Table 4

Comp 11: Upregulated Genes

No. Copies	Description	GenBank Identifier
6	23 kD highly basic protein	HS23KDHBP
103	Act-2	HUMACT2A
1	activated B cell factor 1	AF060154
1	activating transcription factor 3	HUMATF3X
2	adenylyl cyclase-associated protein (CAP)	HUMADCY
47	adipocyte lipid-binding protein	HUMALBP
3	aquaporin 9	AB008775
17	ATPase	HUMH01A
22	B94	HUMB94
1	Cathepsin B	HUMCATHB
10	Cathepsin L	HSCATHL
5	EBV-induced protein	HSU19261
1	Elongation factor 1	HSEF1AC
46	Fibronectin	HSFIB1
57	Guanylate binding protein isoform I (GBP-2)	HUMGBP1
2	IFN-γ inducible early response gene	HSINFGER
33	IGF binding protein 3	HSIGFBP3M

No. Copies	Description	GenBank Identifier
1	IL-1 receptor antagonist	HSI1RA
3	IL-1β	HUMIL1BA
20	IL-8	HSMDNCF
4	Insulin-like growth factor binding protein 3	HSIGFBP3M
3	JKA3 mRNA induced upon T cell stimulation	HSU38443
2	KIAA0251	D87438
3	Macrophage scavenger receptor type I	HUMRMSR1
184	MIP-1α (LD78)	HUMCKLD78
218	MIP-2α (GRO-β)	HUMGROB
50	MIP-2β (GRO-γ)	HUMGROG5
58	Mn SOD	HSMNSOD
4	Musculin	AF087036
1	Paraplegin	HSPARAPLE
3	Prostaglandin endoperoxide synthase-2	HUMPTGS2
1	RANTES	HUMRANTES
1	Reticulocalbin	HUMRCN
1	Ribosomal protein L21	HSU14967
1	Ribosomal protein L7	HSRBPRL7A
1	Ribosomal protein S28	HUMRSPT
34	Spermidine/spermine N1-acetyltransferase	HUMSPERMNA
1	Striated muscle contraction reg. Protein	HUMID2B.
95	TNF-inducible (TSG-6) mRNA	HUMTSG6A
10	TNFα	HSTNFR
2	Translation initiation factor 2β	HUMELF2
1	TRNA-Ala	HSCR6ALAT
17	Type IV collagenase	HUM4COLA
19	EST	AA916304
15	EST	AA873350
10	EST	AA011639
8	EST	AA346072
3	EST	AA284427
2	EST, IL-1/TNF-inducible	HSEST222
2	EST	W88513
1	EST	AA904231
1	EST	AA767777
1	EST	AA969937
1	EST	AA528703
4	Genomic	HSAC000119
4	Genomic	AC000403
1	Genomic	AC004130
6	NO MATCH	
2	NO MATCH	
2	NO MATCH	
1	NO MATCH	
1	NO MATCH	
1	NO MATCH	

No. Copies	Description	GenBank Identifier
1	NO MATCH	
1	NO MATCH	
1157	**Total sequenced (upregulated)**	

Example 7

FACS Analysis of Differentially Expressed Genes from Stimulated and Unstimulated THP-1 Cells

(Experiment: Comp 14)

In a separate experiment, reference DNA population preparation and competitive hybridization were done as described in Example 6. 9150 beads (0.89%), the brightest CY5 off the 1:1 diagonal, were sorted. 11085 beads (1.15%), the brightest R110 off the 1:1 diagonal, were sorted. The identified sequences are listed in Tables 5 and 6.

Table 5

Comp 14: Downregulated Genes

No. Copies	Description
29	H.sapiens mRNA for 23 kD highly basic protein
13	H.sapiens mRNA for ribosomal protein S18
12	Laminin receptor homolog mRNA
12	H.sapiens mRNA for ribosomal protein L26
12	Human ribosomal L5 protein mRNA
9	Human mRNA for elongation factor 1-alpha
9	H.sapiens mRNA for large subunit of ribosomal protein L21
7	H.sapiens gene for ribosomal protein L38
6	Homo sapiens cDNA, 3' end /clone=IMAGE
6	H.sapiens rpS8 gene for ribosomal protein S8
5	Human ribosomal protein L3 mRNA
5	Human Ki nuclear autoantigen mRNA
5	Human ribosomal protein L7a mRNA
5	Novel
5	Human mRNA for ribosomal protein S11
5	Neuroblastoma RAS viral (v-ras) oncogene homolog
5	Human mitochondrial DNA
5	H.sapiens initiation factor 4B cDNA
5	Human endothelial-monocyte activating polypeptide II mRNA
5	Novel
5	Human monocytic leukaemia zinc finger protein (MOZ) mRNA
4	Human platelet activating factor acetylhydrolase, brain isoform, 45 kDa subunit (LIS1) gene

No. Copies	Description
4	Human ferritin L chain mRNA
4	Human DNA sequence from cosmid cN37F10 on chromosome 22q11.2-qter
4	Human mRNA for core I protein
4	H.sapiens mRNA homologous to mouse P21 mRNA
4	H.sapiens mRNA for ribosomal protein L6
4	Human ribosomal protein L9 pseudogene
4	Homo sapiens cDNA, 3' end /clone=486654
4	Human MHC protein homologous to chicken B complex protein mRNA
4	Human elongation factor EF-1-alpha gene
3	H.sapiens Sub1.5 mRNA
3	Human mRNA for Apo1_Human (MER5(Aop1-Mouse)-like protein)
3	Homo sapiens chromosome 5, P1 clone 702A10 (LBNL H56)
3	Human fumarase precursor (FH) mRNA
3	Homo sapiens cDNA, 3' end /clone=IMAGE:1695780
3	Human GST1-Hs mRNA for GTP-binding protein
3	H.sapiens mRNA for RNA polymerase II 140 kDa subunit
3	Homo sapiens ribosomal protein L30 mRNA
3	Human ribosomal protein S17 mRNA
3	HSEST222 Homo sapiens cDNA /clone=MEC-222 /gb=X84721 /gi=673398 /ug=Hs.115716 /len=558
3	Homo sapiens Arp2/3 protein complex subunit p21-Arc (ARC21) mRNA
3	Human cytoplasmic dynein light chain 1 (hdlc1) mRNA
3	Human ribosomal protein S3a mRNA
3	Human mRNA for heat shock protein hsp86
2	Homo sapiens Munc13 mRNA
2	Human translational initiation factor 2 beta subunit (elF-2-beta) mRNA
2	Human mRNA for potential laminin-binding protein
2	Human cyclophilin-related processed pseudogene
2	Homo sapiens ribosomal protein S20 (RPS20) mRNA
2	Human acidic ribosomal phosphoprotein P1 mRNA
2	Human ribosomal protein S13 (RPS13) mRNA
2	Novel
2	Homo sapiens cDNA /clone=IMAGE:979232
2	Homo sapiens cDNA, 3' end /clone=81477
2	Human intercellular adhesion molecule-1 (ICAM-1) mRNA
2	Human mRNA for ribosomal protein L17
2	Human mRNA for carboxyl methyltransferase
2	Human mRNA for cytoskeletal gamma-actin
2	Homo sapiens cDNA, 3' end /clone=626635
2	Human nucleophosmin mRNA
2	Human ribosomal protein L10 mRna
2	Novel
2	Y box binding protein-1 (YB-1) mRNA
2	Human guanylate binding protein isoform I (GBP-2) mRNA
2	Homo sapiens cyclin D3 (CCND3) mRNA
2	Novel
2	Homo sapiens cDNA, 3' end /clone=IMAGE:1474218

No. Copies	Description
2	Homo sapiens Ubiquitin mRNA sequence
2	Human ribosomal protein L7
2	Human phosphotyrosine independent ligand p62 for the Lck SH2 domain mRNA
2	HSC3EF102 Homo sapiens cDNA, 3' end /clone
	Total singlets: 189
	Total contigs: 72
	Total seq reads in contigs: 306
	Total seqs to be searched: 610

Table 6

Comp 14: Upregulated Genes

No. Copies	Description
77	Human mRNA for putative cytokine 21 (HC21)
31	Human gene for tumor necrosis factor (TNF-alpha)
27	Human insulin-like growth factor-binding protein-3 gene
26	Human LD78 alpha gene
23	Human mRNA for macrophage inflammatory protein-2beta (MIP2beta)
20	Human cytokine LD78 gene
20	H.sapiens gene for spermidine/spermine N1-acetyltransferase
20	Human gene for melanoma growth stimulatory activity (MGSA)
17	Human ferritin H chain mRNA
13	Novel
13	Human adipocyte lipid-binding protein
12	Human interleukin 8 (IL8) gene
9	Homo sapiens cDNA, 3' end /clone=73864
8	Human ATL-derived PMA-responsive (APR) peptide mRNA
8	Human ATL-derived PMA-responsive (APR) peptide mRNA
7	Human cell surface glycoprotein CD44 mRNA
7	H.sapiens SOD-2 gene
7	Human hypoxanthine phosphoribosyltransferase (HPRT) gene
6	Human tumor necrosis factor-inducible (TSG-6) mRNA fragment
6	Human adenosine receptor (A2) gene
5	Human phosphatidylinositol 3-kinase catalytic subunit p110delta Mrna
4	Human BAC clone RG104I04 no function
4	Homo sapiens adenosine triphosphatase mRNA
4	Human mRNA (3'-fragment) for (2'-5') oligo A synthetase E
4	Genomic sequence no function
2	Human type IV collagenase mRNA
2	Human ribosomal protein S17 mRNA
2	Homo sapiens cDNA, 3' end /clone=IMAGE:1459553
2	Human interleukin 1-beta (IL1B) gene

Example 8

FACS Analysis of Differentially Expressed Genes from Stimulated and Unstimulated THP-1 Cells
(Experiment: Comp 15)

In a separate experiment, cDNA from stimulated and unstimulated THP-1 cells was prepared for competitive hybridization as described in Example 6. The reference DNA population was prepared as described in Example 6, except that the Comp 15 bead library consisted of 2,570,000 beads, with a complexity of 1 million clones from the THP-1 stimulated library and the THP-1 unstimulated library (50% of each). 13,988 beads (.87%), the brightest CY5 off the 1:1 diagonal, were sorted. 17,393 beads (1.08%), the brightest R110 off the 1:1 diagonal, were sorted. The identified sequences are listed in Tables 7 and 8.

Table 7

Comp 15: Downregulated Genes

No. Copies	Description
25	H.sapiens mRNA for 23 kD highly basic protein
17	Homo sapiens ribosomal protein L30 Mrna
16	H.sapiens mRNA for ribosomal protein S18
15	H.sapiens mRNA for ribosomal protein L6
14	L44-like ribosomal protein (L44L) and FTP3 (FTP3) genes
11	Homo sapiens PYRIN (MEFV) mRNA
9	Human cathepsin G mRNA
8	Human mRNA for ribosomal protein S11
8	Novel
8	H.sapiens mRNA for ribosomal protein L37a
8	Novel
8	H.sapiens mRNA for ribosomal protein L26
8	H.sapiens mRNA for translationally controlled tumor protein p21 Homology
8	Human deoxyuridine triphosphatase (DUT) Mrna
8	Human growth factor independence-1 (Gfi-1) mRNA
7	Human mRNA for ribosomal protein L39
7	Human ribosomal protein L10 mRNA
6	Human ribosomal protein L9 mRNA, complete cds. 5/96
6	Homo sapiens cDNA, 3' end /clone=IMAGE:1862607 /clone_end=3' /gb=AI053436 /ug=Hs.135355 /len=138
6	Human gene for catalase Weak Homology
5	H.sapiens mRNA for ribosomal protein L7
5	Homo sapiens (clone cori-1cl5) S29 ribosomal protein mRNA
5	Human mRNA for HBp15/L22
5	H.sapiens mRNA for NEFA protein
5	Novel
5	Human mRNA for potential laminin-binding protein
5	HSEST222
4	Human ribosomal protein S16 mRNA
4	Homo sapiens cDNA /clone=IMAGE:1118473 /gb=AA603101 /gi=2436962 /ug=Hs.14214 /len=621
4	H.sapiens mRNA for large subunit of ribosomal protein L21
4	Human HMG-17 gene for non-histone chromosomal protein HMG-17
4	Human ribosomal protein L5 mRNA
4	H.sapiens Uba80 mRNA for ubiquitin. 2/97
4	Human interferon-inducible mRNA
3	Homo sapiens mRNA for ribosomal protein L14
3	H.sapiens rpS8 gene for ribosomal protein S8
3	Homo sapiens monocyte/macrophage Ig-related receptor MIR-7 (MIR cl-7) mRNA
3	Homo sapiens U2 snRNP auxiliary factor small subunit
3	Homo sapiens 3-phosphoglycerate dehydrogenase mRNA

No. Copies	Description
3	Novel
3	Homo sapiens aflatoxin aldehyde reductase AFAR mRNA
3	Homo sapiens ribosomal protein L18a mRNA
3	Homo sapiens histone H2A.F/Z variant (H2AV) mRNA
3	Human ribosomal protein L27a mRNA
3	H.sapiens gene for ribosomal protein L38
3	Homo sapiens cDNA /clone=IMAGE:1089890 /gb=AA584384 /gi=2368993 /ug=Hs.100437 /len=434
3	Human ribosomal protein S17 mRNA
3	Human cyclophilin-related processed pseudogene
3	H.sapiens MUC5B gene, rearranged DNA fragment
2	Homo sapiens gene for ribosomal protein L41
2	Homo sapiens glia maturation factor beta mRNA
2	Novel
2	Human ribosomal protein L7a
2	Human ribosomal protein S13 (RPS13) mRNA
2	Homo sapiens cDNA /clone=IMAGE:979448 /gb=AA523303 /gi=2264015 /ug=Hs.15476 /len=640
2	Human profilin mRNA
2	Homo sapiens cDNA, 3' end /clone=1391189 /clone_end=3' /gb=AA781132 /ug=Hs.110803 /len=658
2	Human mRNA for mitochondrial ATP synthase (F1-ATPase) alpha subunit
2	Human mRNA for cytoskeletal gamma-actin
2	Human mRNA for ribosomal protein L32
2	H.sapiens beta-sarcoglycan gene
2	Human mRNA for 26S proteasome subunit p31
2	H.sapiens mRNA for ribosomal protein S15a
2	Novel
2	Novel
2	Homo sapiens IgE receptor beta chain (HTm4) mRNA
2	Human HuR RNA binding protein (HuR) mRNA
2	human alpha-tubulin mRNA
2	H.sapiens mRNA for elongations factor Tu-mitochondrial
2	Homo sapiens cDNA, 3' end /clone=550365 /clone_end=3' /gb=AA098869 /gi=1644973 /ug=Hs.103088 /len=526
2	Homo sapiens cDNA, 3' end /clone=448402 /clone_end=3' /gb=AA777529 /ug=Hs.11355 /len=529
2	Human mRNA for proteasome subunit HsC10-II
2	Homo sapiens RCL (Rcl) mRNA
2	Homo sapiens clone DT1P1A10 mRNA, CAG
2	Novel
2	Human prothymosin-alpha gene
	Total singlets: 213
	Total contigs: 76
	Total seq reads in contigs: 366
	Total seqs to be searched: 717

Table 8

Comp 15: Upregulated Genes

No. Copies	Description
188	Human gene for tumor necrosis factor (TNF-alpha)
61	Cytochrome P450
38	H.sapiens mRNA for uridine phosphorylase
27	HuEST
14	Human tumor necrosis factor-inducible (TSG-6) mRNA fragment
11	Homo Sapiens Chromosone 21 clone
8	Novel
6	Novel
6	Novel
5	SOD-2 Gene
5	Human LD78 beta gene
4	Adenosine receptor A2
3	Human Mitocondrial DNA
3	Human spermidine/spermine N1-acetyltransferase (SSAT) gene
3	H.sapiens mRNA for 23 kD highly basic protein
3	HuEST
3	Human tumor necrosis factor alpha inducible protein A20 mRNA
2	Human spermidine/spermine N1-acetyltransferase (SSAT) gene
2	Human plasma membrane Ca2+ pumping ATPase mRNA
2	Cathepsin L
2	GRO3 oncogene MIP2-beta
2	Small inducible cytokine A4 (homologous to mouse Mip-1b) ACT2
2	GRO2 oncogene MIP2-alpha
2	Human LD78 alpha gene
2	Interleukin 8
	Total singlets: 91
	Total contigs: 25
	Total seq reads in contigs: 404
	Total seqs to be searched: 726

Example 9

Isolation of Rare Genes From Stimulated THP-1 Cells

(Experiment: Cot 3)

In this example, rare genes are isolated from stimulated THP-1 cells by collecting beads of lower relative intensity. Bead and probe libraries were constructed from mRNA prepared from phorbol ester treated THP-1 cultured cells. Six bead libraries (160K complexity) were loaded twice to BP 11 combitagged beads. A total of 1,260,000 beads were sorted. The beads were filled in and ligated. The top

strand of the beads was stripped with 2.5 ml 150 mM NaOH washes at room temperature for 15 minutes with mild vortexing. The beads were washed twice in 0.5 ml of 4X SSC/0.1% SDS. 100,000 beads were hybridized overnight with 50 ng of CY5 labelled probe from stimulated THP-1 cells in 4X SSC/0.1% SDS at 65°. The recovered samples were rinsed 2 times with 1X SSC/0.1% SDS, resuspended in 0.5 ml of 1X SSC/0.1% SDS, and washed at 65°C for 15 minutes. The beads were then rinsed in 0.1X SSC/0.1% SDS and washed at 55°C in 0.1X SSC/0.1% SDS for 15 minutes. 98,880 clones were analyzed and sorted by flow cytometry. Sample CT003E contained 126 clones which barely hybridized any CY5 probe. Sample CT003F contained 1557 clones that did not find enough probe to migrate to the diagonal. These beads contained the least frequent copies in our probe library. 50 clones from each gate (see Figure 7) were picked for sequence analysis. The identified sequences are listed in Table 9.

Table 9

THP-1 Rare Genes

No. Copies	Descripton	GenBank Identifier
CT003E		
2	Alu primary transcript	U67828
1	AMP deaminase	HSAMPD3B
1	BBC1	HSBBC1
14	CD44	HUMCD44B
1	clone 23933 mRNA	HSU79273
7	EST	AA905212
1	EST	AA975736
1	EST	N53143
1	EST	AA808221
1	EST	AA826047
1	EST	AA736779
1	EST	AA994497
1	EST	AI049999
1	EST (88% homology)	AA626040
8	EST (contains Alu repeat)	AA129219
9	EST (contains Alu repeat)	AI085719
1	EST (contains Alu repeat)	W07654
1	EST (Sau3A not present)	AA553627
3	ferritin H chain	HUMFERH
1	IL1 β	HUMIL1BA

No. Copies	Descripton	GenBank Identifier
1	KIAA0098 gene mRNA	HUMKG1DD
1	mito. cyt oxidase subunit I pseudogene	AF035429
1	Mitochondrial genome	MIHSGENOM
1	NADH:ubiquinone oxidoreductase NDUF56 subunit	AF044959
2	no match	
1	no match	
1	only 12 bases	
1	only 12 bases	
1	only 13 bases	
1	Pyruvate kinase, M gene for M1-type & M2-type	HSPKM12
4	TNF α	HSTNFR
7	TNF type I recept. assoc. prot./DNAse I/HSP75	HSU12595/D83195/AF043254
1	type IV collagenase	HUM4COLA
1	Ubiquitin hydrolyzing enzyme I (UBHI)	AF022789
1	VASP gene	HSVASP413
CT003F		
1	Apolipoprotein C-II	HSAPOC2G
3	BBC1	HSBBC1
1	clone s153 mRNA fragment	HUMFRCC
5	cytoskeletaλ γ actin	HSACTCGR
1	elongation factor 1 α	HSEF1AC
1	EST	AA905212
2	EST	AA977353
1	EST	AA135810
1	EST (contains Alu repeat)	H08741
1	EST	AA282788
1	EST	AA226660
1	EST (85% homology; contains Alu, CACA tract)	AA704393
1	EST (86% homology; contains Alu)	H60533
1	EST (88% homology)	AA228701
2	EST (contains Alu repeat)	AI085719
1	EST (contains Alu repeat)	AA713891
1	EST (rat)	AI136745
1	ferritin H chain	HUMFERH
1	genomic (72 bp; 88% homology)	HSAC002082
1	ICAM-1	HUMICAMA1M
1	IL1 β	HUMIL1BA
1	Interferon γ receptor accessory factor 1	HSU05877
2	mito. cyt oxidase subunit I pseudogene	AF035429
1	no match	
1	no match	

No. Copies	Descripton	GenBank Identifier
1	no match (40 bp)	
1	OTK27	D50420
1	p65, rat (partial match to hu synaptotagmin I)	RRP65
1	rat α tubulin (100% hom to rat; 78% to human)	RNATUBZ
1	Ribophorin II	HSRIBIIR
3	Ribosomal protein S3	HSU14991
1	RING4	HSRING4
12	sec61-complex β subunit	HUMSEC61B
1	TNF α	HSTNFR
1	type IV collagenase	HUM4COLA

Example 10

Isolation of Rare Genes From Human Bone Marrow

Bead and probe libraries were constructed from commercially available mRNA from bone marrow. Six bead libraries (160K complexity) were loaded twice to BP 12 combitagged beads. They formed mixes 216, 217, 218, and 219. A total of 3,150,000 beads were sorted. The beads were filled in and ligated. The top strand of mix 217 was stripped off with NaOH. The CT1 bone marrow probe was linearly amplified with CY5 nucleotides and then purified. 200,000 beads were hybridized with 5 and 50 ng of probe overnight at 65°. 180,000 clones from the 5 nG hybridization were interrogated and sorted. Sample CT001 contained 996 clones which barely hybridized any CY5 probe. CT002 sample contained 1988 clones that did not find enough probe to migrate to the diagonal. These beads contained the least frequent copies in our probe library. 200 clones from each gate (see Figure 8) were picked for sequence analysis.

Example 11

FACS Analysis of Differentially Expressed Genes from Normal and Glucose Starved Human Muscle Tissue

Bead and probe libraries were constructed from mRNA prepared from muscle tissue in two states: glucose normal (basal) and glucose starved (clamp). Six bead libraries (160K complexity) from the glucose normal state were loaded to BP 12 combitagged beads to form mix 237. A total of 810,000 beads were sorted. The beads were filled in and ligated. The beads were digested with DpnII enzyme and ligated to an adapter with FITC on the strand opposite to the covalently attached DNA

strand. The top strand of mix 217 was stripped off with NaOH. The CT1 glucose normal probe (13,510,000 complexity) was linearly amplified with CY5 nucleotides and then purified. The CT2 glucose starred probe (7,132,000 complexity) was linearly amplified with R110 nucleotides and then purified. 250,000 beads were hybridized with 5ug of each probe overnight at 65°. 230,000 clones were interrogated and sorted. Sample UP001 contained 968 clones which were upregulated. Sample DN001 contained 1652 clones which were down regulated. 1000 clones from each gate (see Figure 9) were picked for sequence analysis. The identified sequences are listed in Tables 10 and 11.

Table 10

Downregulated Genes in Starved Human Muscle

No. Copies	Description
23	Human mRNA for slow skeletal troponin C
27	Human alkali myosin light chain 1 Mrna
14	Human messenger RNA for beta-globin
13	Human lymphocytic antigen CD59/MEM43 mRNA
10	P6=cytochrome c oxidase subunit Vlc homolog
8	H.sapiens mRNA homologous to mouse P21 mRNA
5	Human SPARC/osteonectin mRNA
5	3', mRNA sequence
4	Pan troglodytes beta-2-microglobulin mRNA
4	reductase
4	Homo sapiens gene for ribosomal protein L41
3	Homo sapiens ribosomal protein L30 mRNA
3	IMAGE: 1388067
2	ni65c01.s1 NCl_CGAP_Pr12 Homo sapiens cDNA clone IMAGE:981696
2	Homo sapiens mRNA for KIAA0454 protein
2	Human gene for cardiac beta myosin heavy chain

Table 11

Upregulated Genes in Starved Human Muscle

No. Copies	Description
4	Human mitochondrion cytochrome b gene
4	Homo sapiens sarcosin mRNA
4	laminin receptor homolog
3	H.sapiens mRNA for 23 kD highly basic protein
3	Human ENO3 mRNA for beta-enolase
3	alpha-tropomyosin
3	alpha B-crystallin
3	Human mRNA for muscle phosphofructokinase
2	Baboon beta-myosin heavy-chain mRNA
2	Human mRNA 3'-fragment for glycogen phosphorylase
2	Human ribosomal L5 protein mRNA
2	H.sapiens mRNA for ribosomal protein L37a
2	Human cytochrome c oxidase subunit VII (COX8) mRNA

All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

We claim:

1. A method of analyzing differential gene expression, comprising:

providing a reference population of nucleic acid sequences attached to separate solid phase supports in clonal subpopulations;

providing a population of polynucleotides of expressed genes from a first cell or tissue source and at least one population of polynucleotides of expressed genes from a different cell or tissue source, the polynucleotides of expressed genes from each source comprising a light-generating label different from the label comprised by polynucleotides of any other source;

competitively hybridizing the populations of polynucleotides of expressed genes from each source with the reference nucleic acid population to form duplexes between the nucleic acid sequences of the reference nucleic acid population and the polynucleotides of each source such that the polynucleotides are present in duplexes on each of the solid phase supports in ratios directly related to the relative expression of their corresponding genes in the sources; and

detecting a relative optical signal generated by the light-generating labels of the duplexes attached thereto.

2. The method of Claim 1, wherein said nucleic acid sequences are DNA sequences.

3. The method of Claim 2, wherein said step of providing said reference population further includes:

forming at least one population of tag-cDNA conjugates from mRNA extracted from at least one of said sources and a repertoire of oligonucleotide tag;

removing a sample of the tag-cDNA conjugates; and

amplifying the tag-cDNA conjugates of the sample.

4. The method of Claim 3, wherein said populations of tag-cDNA conjugates are formed from mRNA extracted from each of said sources, the method further comprising combining said populations of tag-cDNA conjugates from each of said sources prior to removing said sample.

5. The method of Claim 4, wherein said sample is sufficiently small relative to said total tag-cDNA conjugates that substantially all different cDNAs have different oligonucleotide tags.

6. The method of Claim 5, wherein said step of providing said reference population further includes attaching said tag-cDNA conjugates of said sample to said separate solid phase supports by specifically hybridizing said oligonucleotide tags of said tag-cDNA conjugates to their respective complements.

7. The method of Claim 6, wherein said step of amplifying comprises replicating said tag-cDNA conjugates of said sample in a polymerase chain reaction.

8. The method of Claim 6, wherein said step of amplifying comprises replicating said tag-cDNA conjugates of said sample by inserting said tag-cDNA conjugates into a cloning vector and transfecting a host cell therewith.

9. The method of Claim 6, wherein said sample includes a number of oligonucleotide tags less than or equal to one percent of said oligonucleotide tags in said repertoire.

10. The method of Claim 2, wherein said reference DNA population is derived from said expressed genes of all of said sources being analyzed.

11. The method of Claim 2, further comprising sorting each solid phase support according to said relative optical signal.

12. The method of Claim 2, wherein said different light-generating labels are different fluorescent labels.

13. The method of Claim 12, wherein said population of polynucleotides of expressed genes are populations of cDNAs.

14. The method of Claim 13, further comprising the steps of:

accumulating each said solid phase support having said relative optical signal with a value within one or more predetermined ranges of values corresponding to a difference in gene expression among said sources; and

identifying said polynucleotides on each of said solid supports by determining a nucleotide sequence of a portion of each of said polynucleotides.

15. The method of Claim 14, wherein said relative optical signal is a ratio of fluorescence intensities and wherein said populations of polynucleotides are from two sources.

16. The method of Claim 15, wherein said portion of said polynucleotides is a sequence of at least ten nucleotides.

17. The method of Claim 15, wherein said step of identifying includes simultaneous sequencing of at least ten thousand of said polynucleotides by massively parallel signature sequencing.

18. A method of isolating polynucleotides derived from genes differentially expressed in a plurality of different cells or tissues, the method comprising the steps of:

providing a reference DNA population of DNA sequences attached to separate microparticles in clonal subpopulations;

providing a population of polynucleotides derived from genes expressed in each of the plurality of different cells or tissues, each polynucleotide having a light-generating label capable of generating an optical signal indicative of the cells or tissues from which it is derived;

competitively hybridizing the populations of polynucleotides of genes expressed in each of the plurality of different cells or tissues with the reference DNA population to form duplexes between the DNA sequences of the reference DNA population and polynucleotides from each of the different cells or tissues such that the

polynucleotides are present in duplexes on each of the microparticles in ratios directly related to the relative expression of their corresponding genes in the different cells or tissues; and

isolating polynucleotides corresponding to genes differentially expressed in the different cells or tissues by sorting microparticles in accordance with the optical signals generated by the populations of polynucleotides hybridized thereto.

19. The method of Claim 18, wherein said reference DNA population is derived from genes expressed in the plurality of different cells or tissues being analyzed.

20. The method of Claim 19, wherein said plurality of different cells or tissues is two and wherein said optical signal is a fluorescent signal.

21. The method of Claim 20, wherein said populations of polynucleotides are labeled with different fluorescent labels.

22. The method of Claim 21, wherein said populations of polynucleotides are populations of cDNAs.

23. The method of Claim 22, wherein said step of competitively hybridizing includes providing hybridization conditions which result in substantially all of said duplexes being perfectly matched duplexes.

24. The method of Claim 23, wherein said step of isolating includes sorting said microparticles in accordance with the ratio of fluorescence intensities generated by said populations of cDNAs hybridized thereto.

25. The method of Claim 24, wherein said step of isolating includes sorting said microparticle with a fluorescence-activated cell sorter.

26. The method of Claim 25, further including the step of identifying said isolated cDNAs by determining a nucleotide sequence of a portion of each said isolated cDNA.

27. A method of determining relative abundance of gene products, comprising:

providing a reference DNA population of DNA sequences attached to separate solid phase supports in clonal subpopulations;

providing a population of polynucleotides derived from genes expressed in at least one cell or tissue source, the polynucleotides having a light-generating label;

hybridizing the polynucleotides with the reference DNA population to form duplexes between the DNA sequences of the reference DNA population and the polynucleotides; and

sorting each solid phase support according to the optical signal generated by the light-generating labels of the duplexes attached thereto,

wherein relative abundance of the gene products is correlated with the relative level of intensity of the optical signals obtained from the duplexes, wherein a lower intensity is indicative of a rarer gene product.

28. The method of Claim 27, further comprising isolating solid phase supports having lower relative intensities, wherein said isolated solid phase supports comprise at most about 5% of the total solid phase supports provided.

29. The method of Claim 28, wherein said isolated solid phase supports comprise at most about 0.5% of the total supports provided.

30. A method of isolating polynucleotides according to the abundance of the nucleic acid sequences from which they are derived, comprising:

providing a reference DNA population of DNA sequences attached to separate microparticles in clonal subpopulations;

providing a population of polynucleotides derived from nucleic acid sequences present in the cells of at least one cell or tissue source, each polynucleotide having a light-generating label capable of generating an optical signal;

competitively hybridizing the population of polynucleotides with the reference DNA population to form duplexes between the DNA sequences of the reference DNA population and the polynucleotides, the hybridizing being conducted under conditions which provide a hybridization rate proportionate to the abundance of the polynucleotide wherein less abundant polynucleotides would remain unhybridized;

sorting the polynucleotides into a hybridized population and an unhybridized population.

31. The method of Claim 30, wherein said polynucleotides are hybridized with said reference DNA population under conditions such that said unhybridized population comprises polynucleotides derived from rare gene products.

32. The method of Claim 30, wherein said polynucleotides are hybridized with said reference DNA population under conditions such that said unhybridized population is substantially enriched in polynucleotides derived from nonrepetitive nucleic acid sequences.

33. A composition comprising a mixture of microparticles, each microparticle having a population of identical single stranded nucleic acid molecules attached thereto, the single stranded nucleic acid molecules being different on each microparticle and comprising an oligonucleotide tag in juxtaposition with a polynucleotide derived from an mRNA of at least one cell or tissue source.

34. The composition of Claim 33, wherein said nucleic acid molecules are DNA.

35. The composition of Claim 34, wherein said polynucleotides are derived from a plurality of cell or tissue sources.

36. The composition of Claim 35, wherein said mixture comprises at least 100 different microparticles.

37. The composition of Claim 35, wherein said mixture comprises at least 1000 different microparticles.

38. The composition of Claim 35, wherein said mixture comprises at least 10^4 different microparticles.

39. The composition of Claim 35, wherein said oligonucleotide tag is about 12 to about 60 nucleotides in length.

40. The composition of Claim 35, wherein said oligonucleotide tag is about 18 to about 40 nucleotides in length.

41. The composition of Claim 35, wherein said oligonucleotide tag is about 25 to about 40 nucleotides in length.

42. A composition comprising a mixture of microparticles, each microparticle having a population of identical single stranded nucleic acid molecules attached thereto, the single stranded nucleic acid molecules being different on each microparticle and each of the different nucleic acid molecules comprising a polynucleotide encoding a protein selected from the group consisting of cell cycle proteins, signal transduction pathway proteins, oncogene gene products, tumor suppressors, kinases, phosphatases, transcription factors, growth factor receptors, growth factors, extracellular matrix proteins, proteases, cytoskeletal proteins, membrane receptors, Rb pathway proteins, p53 pathway proteins, proteins involved in metabolism, proteins involved in cellular responses to stress, cytokines, proteins involved in DNA damage and repair, and proteins involved in apoptosis.

43. The composition of Claim 42, wherein each of said nucleic acid molecules further comprises an oligonucleotide tag in juxtaposition with said polynucleotide and positioned between said microparticle and said polynucleotide.

44. The composition of claim 43, wherein each of said microparticles comprises a set of oligonucleotide tags having a sequence different from the oligonucleotide tags of any other microparticle in said composition.

45. The composition of Claim 42, wherein said polynucleotides encode kinases.

46. The composition of Claim 42, wherein said polynucleotides encode cell-cycle proteins.

47. The composition of Claim 42, wherein said polynucleotides encode signal transduction pathway proteins.

48. The composition of Claim 42, wherein said polynucleotides encode proteins involved in apoptosis.

49. The composition of Claim 42, wherein said polynucleotides encode proteins involved in metabolism.

50. A kit for preparing a reference population, comprising:

a plurality of microparticles having oligonucleotide tag complements attached thereto, the oligonucleotide tag complement sequence being different on each microparticle.

51. The kit of Claim 50, further comprising a plurality of vectors comprising a library of tags, the tags having sequences complementary to said tag complements.

52. The kit of Claim 51, further comprising a population of polynucleotides from at least one cell or tissue source.

53. The kit of Claim 52, wherein said polynucleotides are cDNAs.

54. The kit of Claim 52, wherein said population of polynucleotides is contained in a container separate from said plurality of microparticles.

55. The kit of Claim 51, further comprising at least one reagent for preparing said reference population.

56. A kit for analyzing differentially expressed genes, comprising:

a mixture of microparticles, each microparticle having a population of identical single stranded nucleic acid molecules attached thereto, the single stranded nucleic acid molecules being different on each microparticle and comprising polynucleotide derived from an mRNA of at least one cell or tissue source.

57. The kit of Claim 56, wherein each of said nucleic acid molecules further comprises an oligonucleotide tag in juxtaposition with said polynucleotide and positioned between said microparticle and said polynucleotide.

58. The kit of Claim 56, further comprising printed instructions for use in analyzing differentially expressed genes.

59. The kit of Claim 56, further comprising a container.

60. The kit of Claim 56, further comprising a population of cDNA molecules from at least one of said cell or tissue sources.

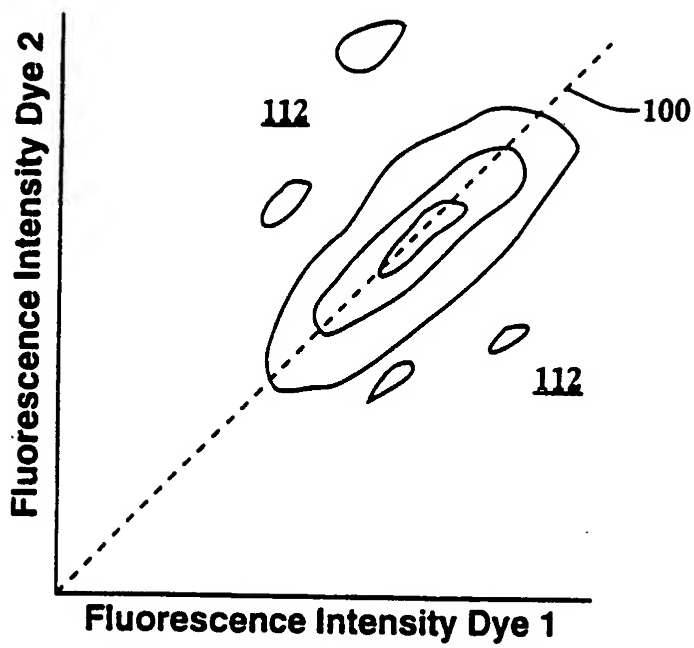

112
100
112
Fluorescence Intensity Dye 2
Fluorescence Intensity Dye 1

Fig. 1A

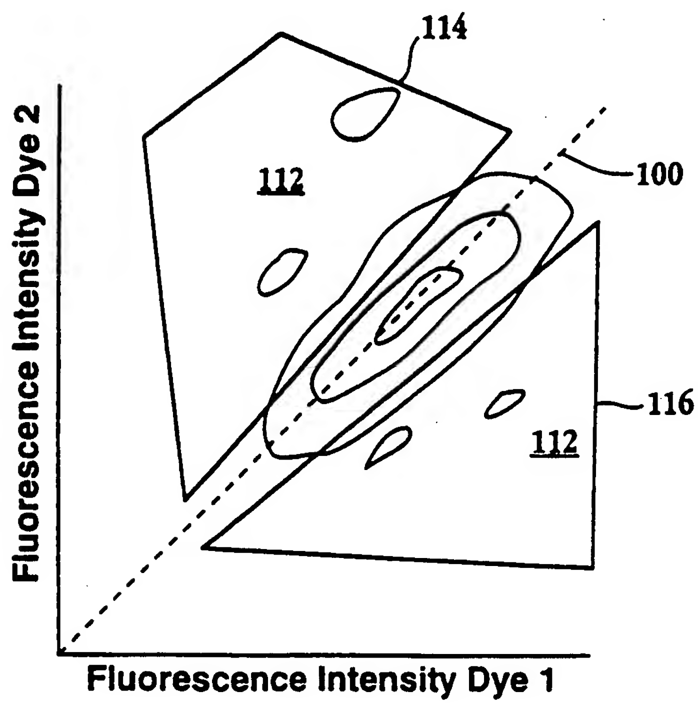

114
112
100
116
112
Fluorescence Intensity Dye 2
Fluorescence Intensity Dye 1

Fig. 1B

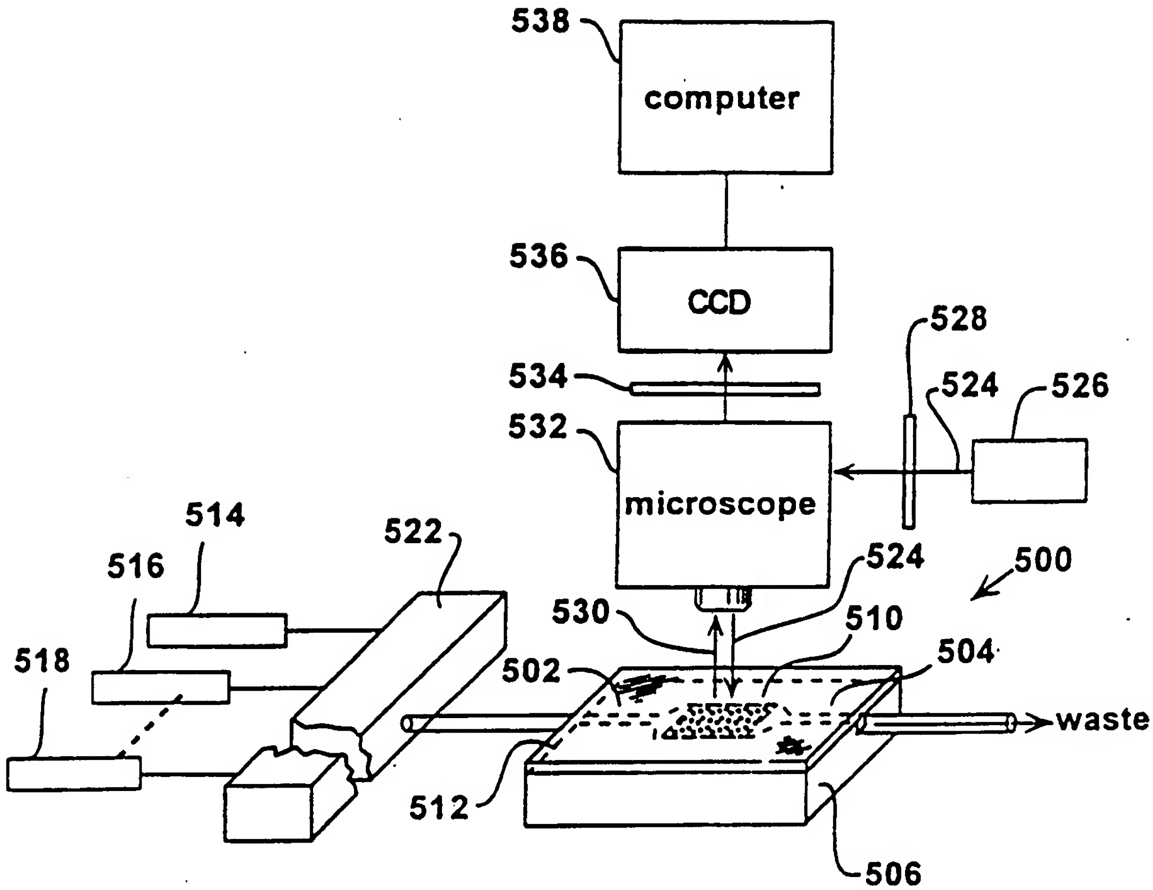

538
computer
536
CCD
528
534
524
526
532
microscope
514
522
516
524
500
530
510
518
502
504
waste
512
506

Fig. 2

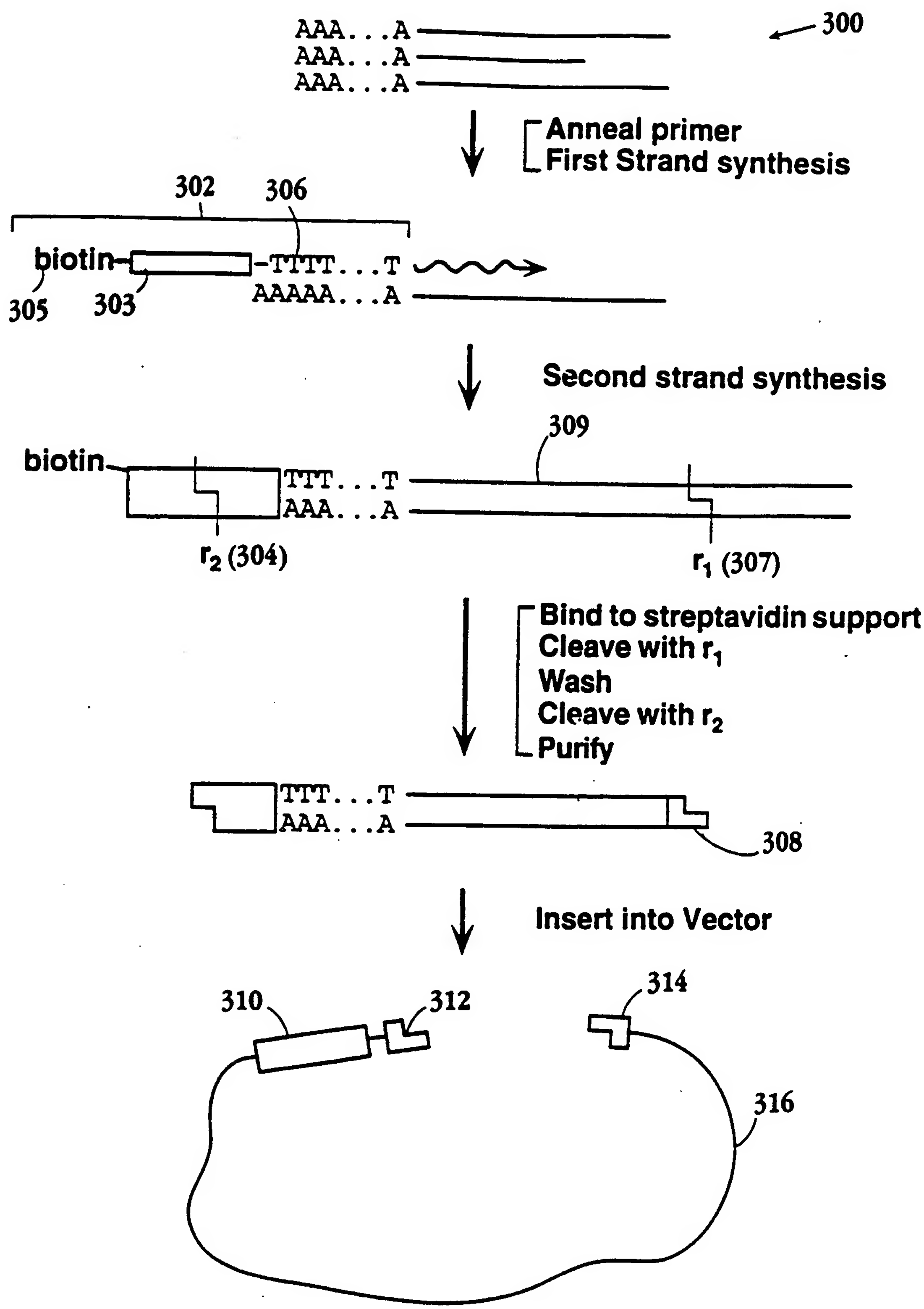

300
AAA...A
AAA...A
AAA...A
Anneal primer
First Strand synthesis
302
306
biotin
-TTTT...T
AAAAA...A
305
303
Second strand synthesis
309
biotin
TTT...T
AAA...A
r_2 (304)
r_1 (307)
Bind to streptavidin support
Cleave with r_1
Wash
Cleave with r_2
Purify
TTT...T
AAA...A
308
Insert into Vector
310
312
314
316

Fig. 3A

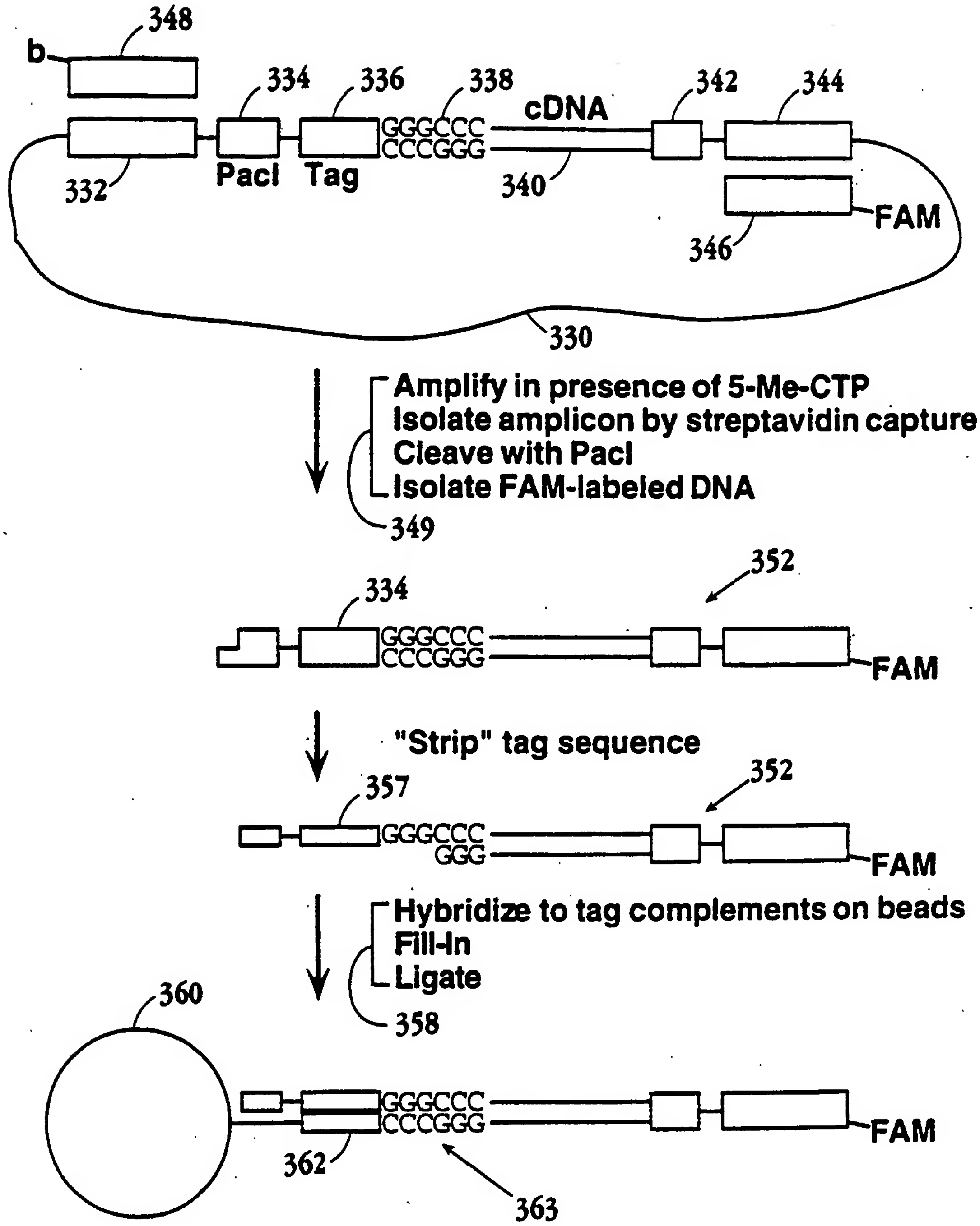

348
b
334
336
338
cDNA
342
344
GGGCCC
CCCGGG
332
PacI
Tag
340
FAM
346
330
Amplify in presence of 5-Me-CTP
Isolate amplicon by streptavidin capture
Cleave with PacI
Isolate FAM-labeled DNA
349
352
334
GGGCCC
CCCGGG
FAM
"Strip" tag sequence
357
352
GGGCCC
GGG
FAM
Hybridize to tag complements on beads
Fill-In
Ligate
358
360
GGGCCC
CCCGGG
FAM
362
363

Fig. 3B

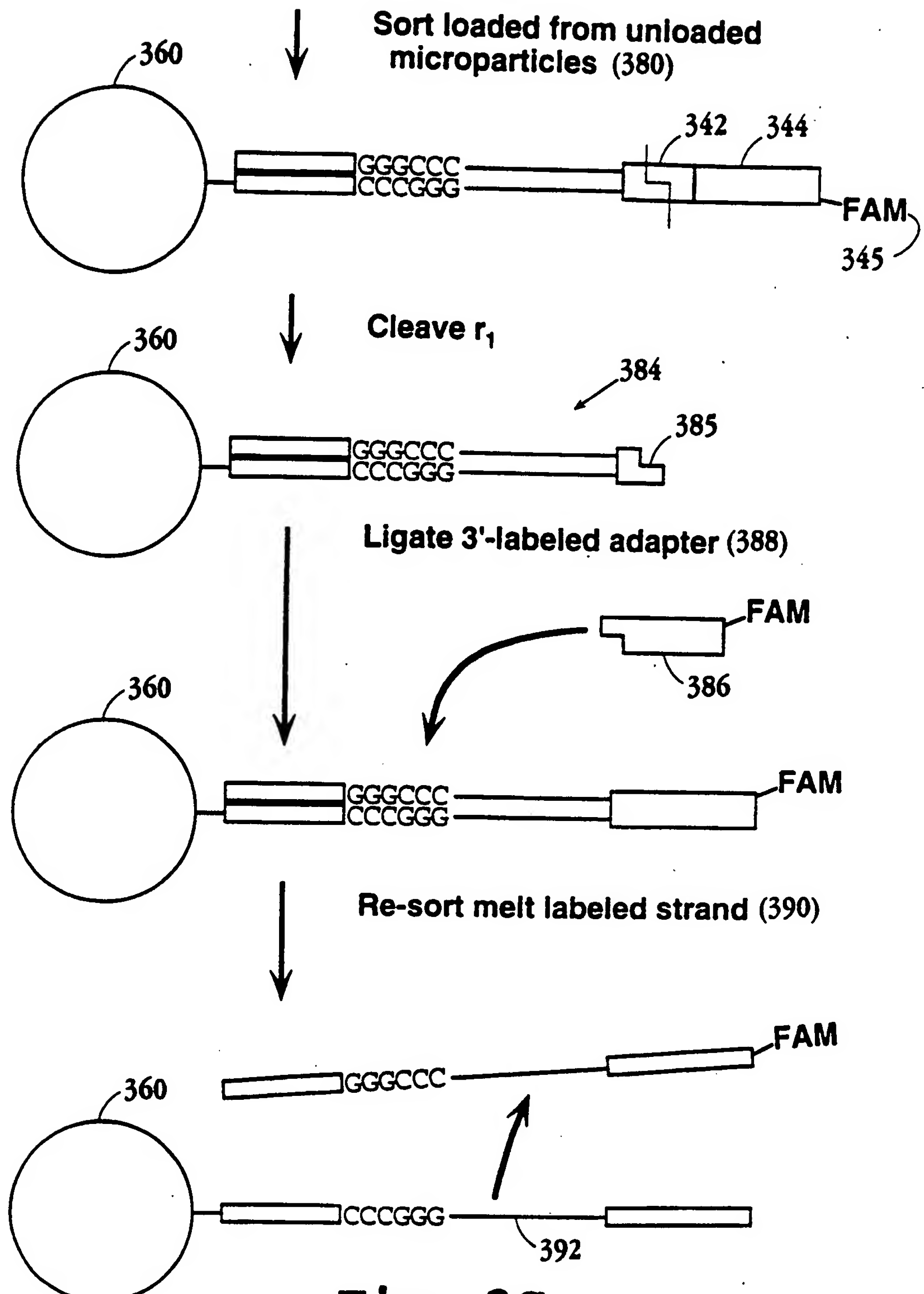

Sort loaded from unloaded microparticles (380)
360
342
344
GGGCCC
CCCGGG
FAM
345
Cleave r_1
360
384
385
GGGCCC
CCCGGG
Ligate 3'-labeled adapter (388)
FAM
386
360
GGGCCC
CCCGGG
FAM
Re-sort melt labeled strand (390)
FAM
360
GGGCCC
CCCGGG
392

Fig. 3C

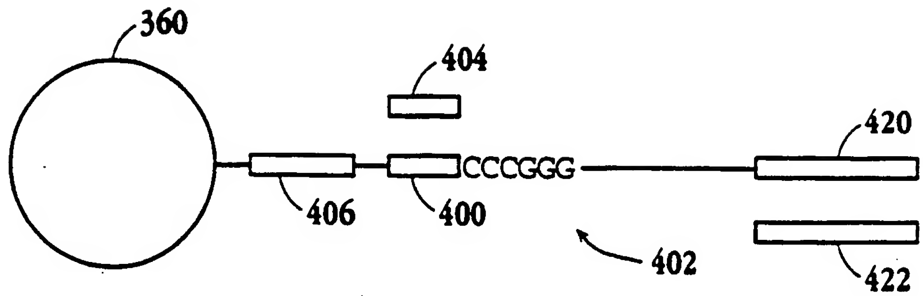

360
404
420
CCCGGG
406
400
402
422

Fig. 4A

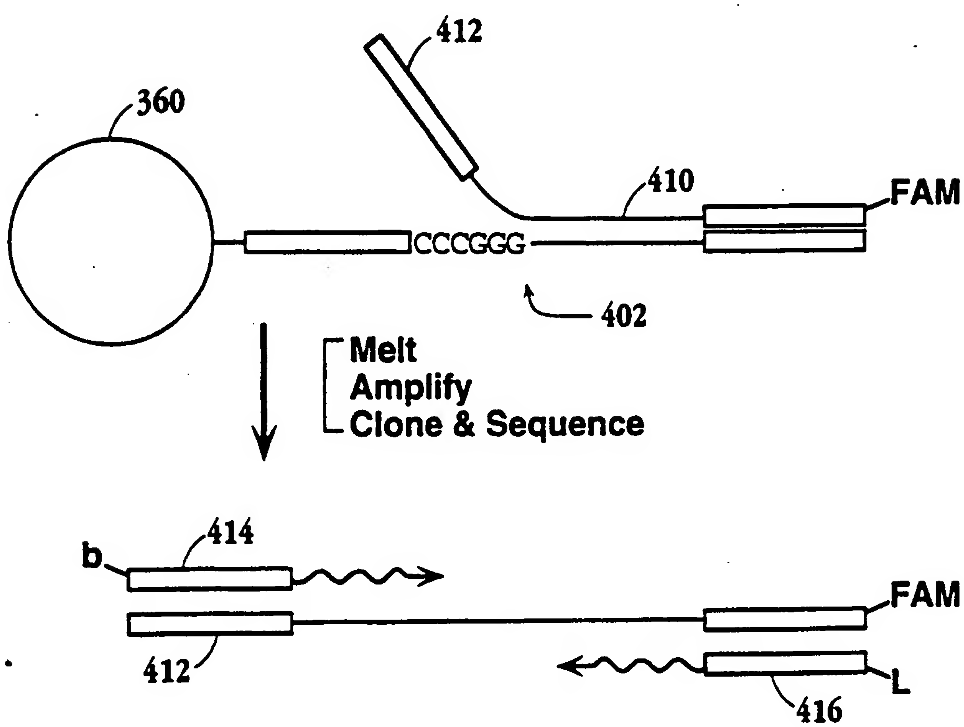

412
360
410
FAM
CCCGGG
402
Melt
Amplify
Clone & Sequence
414
b
FAM
412
L
416

Fig. 4B

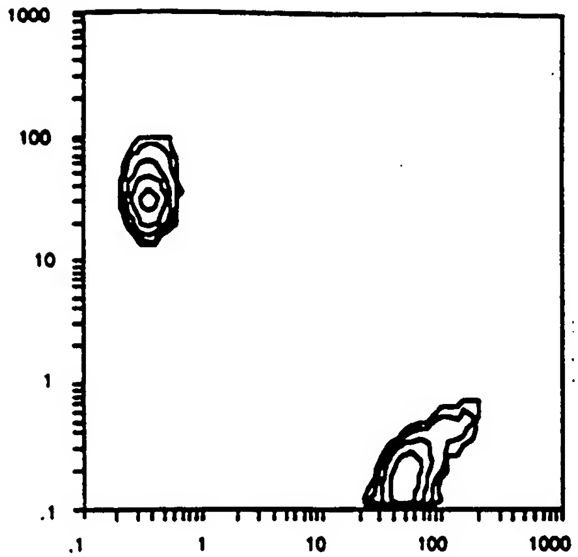


Fig. 5A

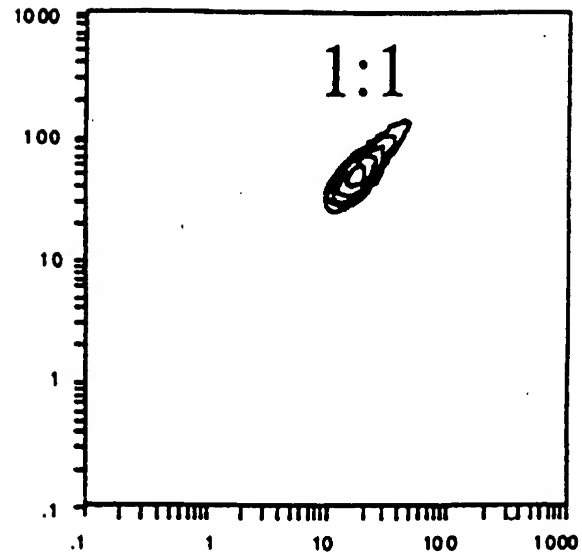


Fig. 5B

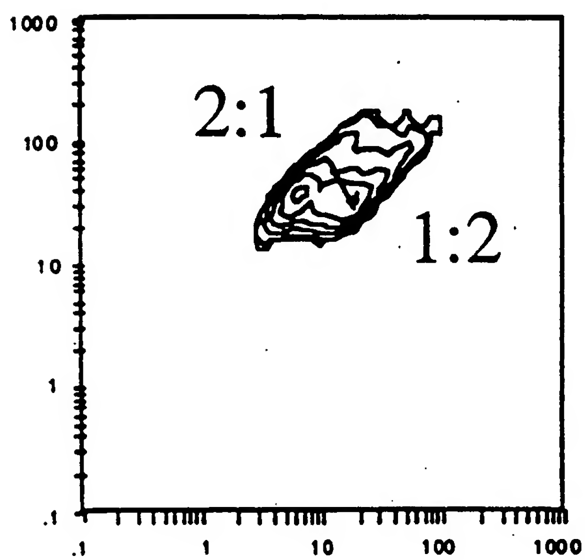


Fig. 5C

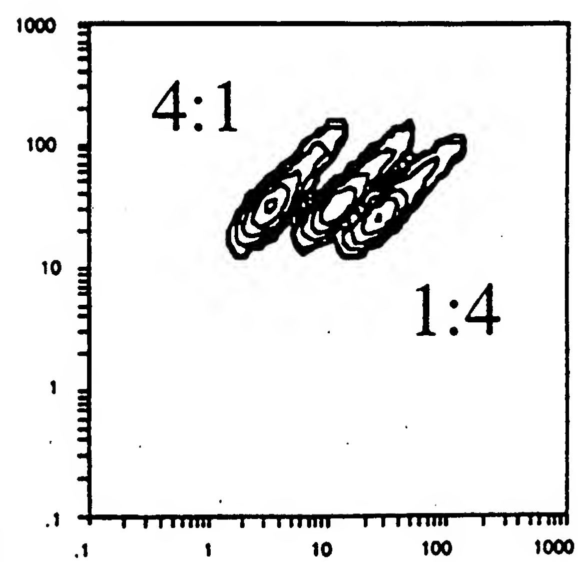


Fig. 5D

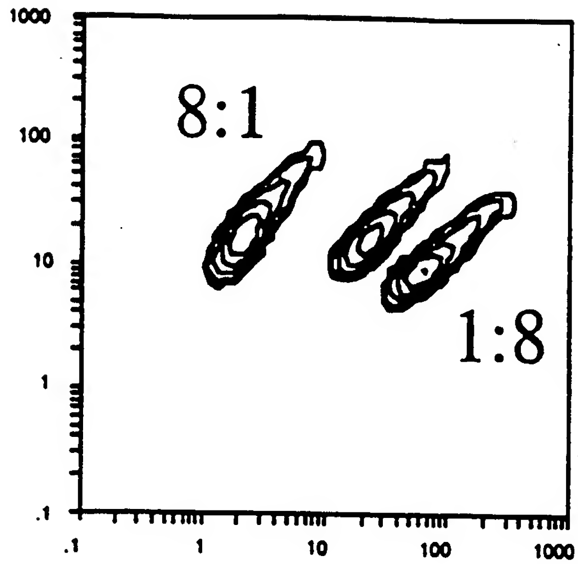


Fig. 5E

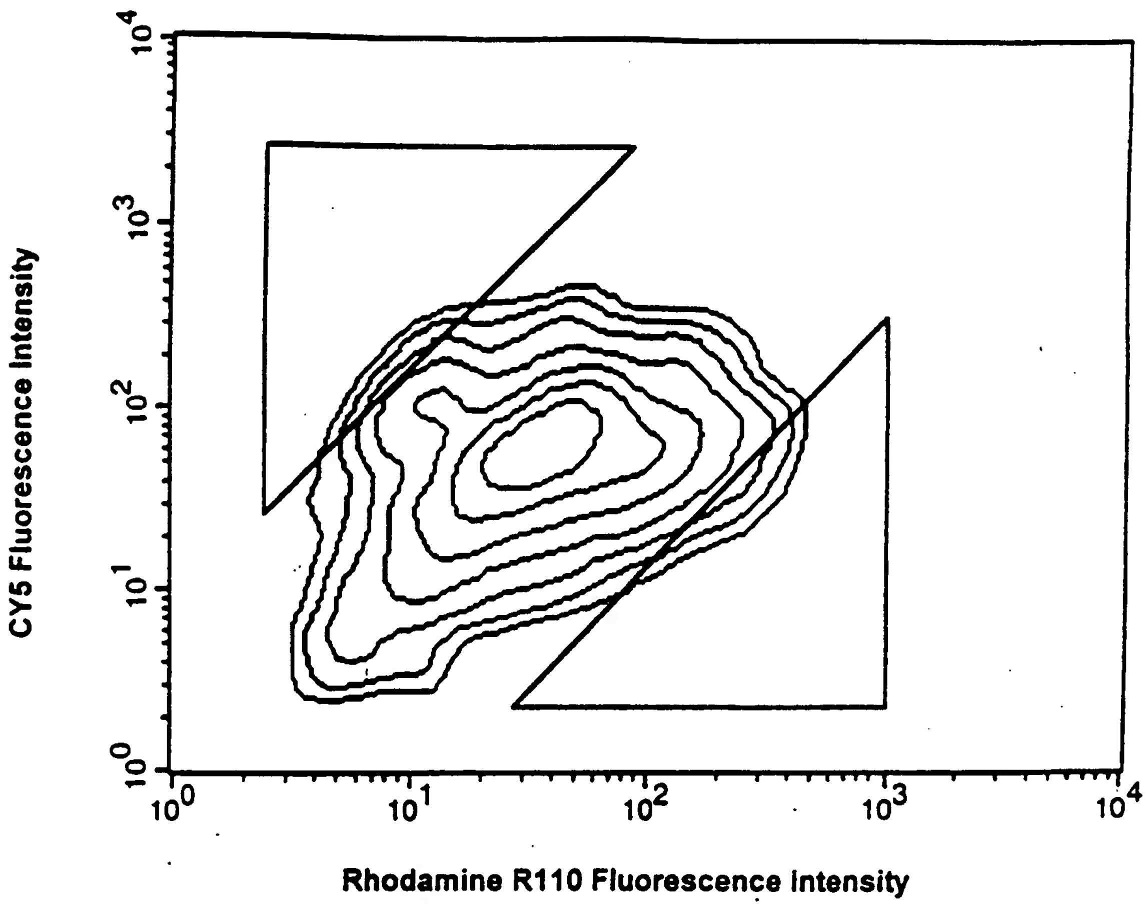

CY5 Fluorescence Intensity
10^0
10^1
10^2
10^3
10^4
Rhodamine R110 Fluorescence Intensity

Fig. 6

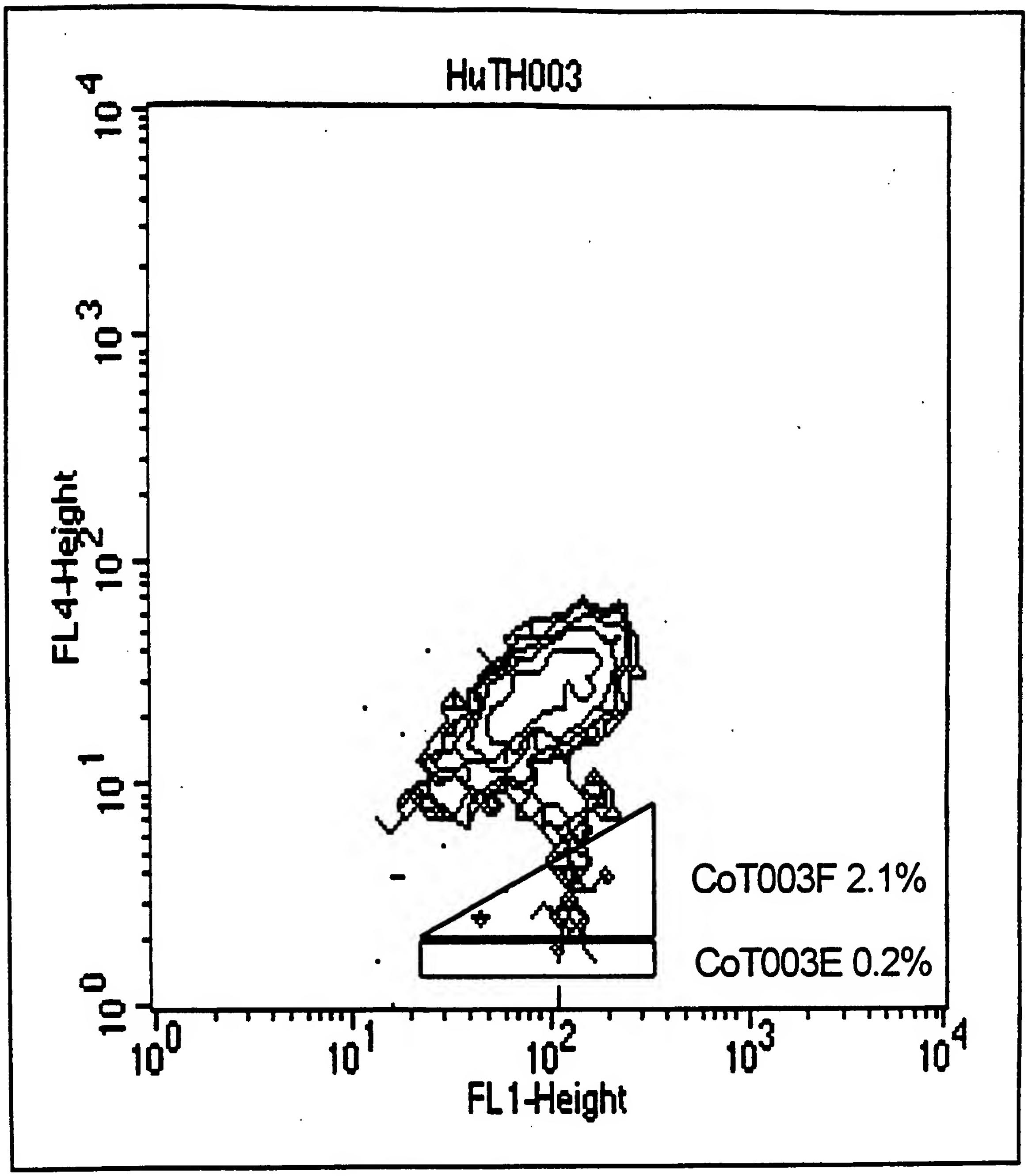
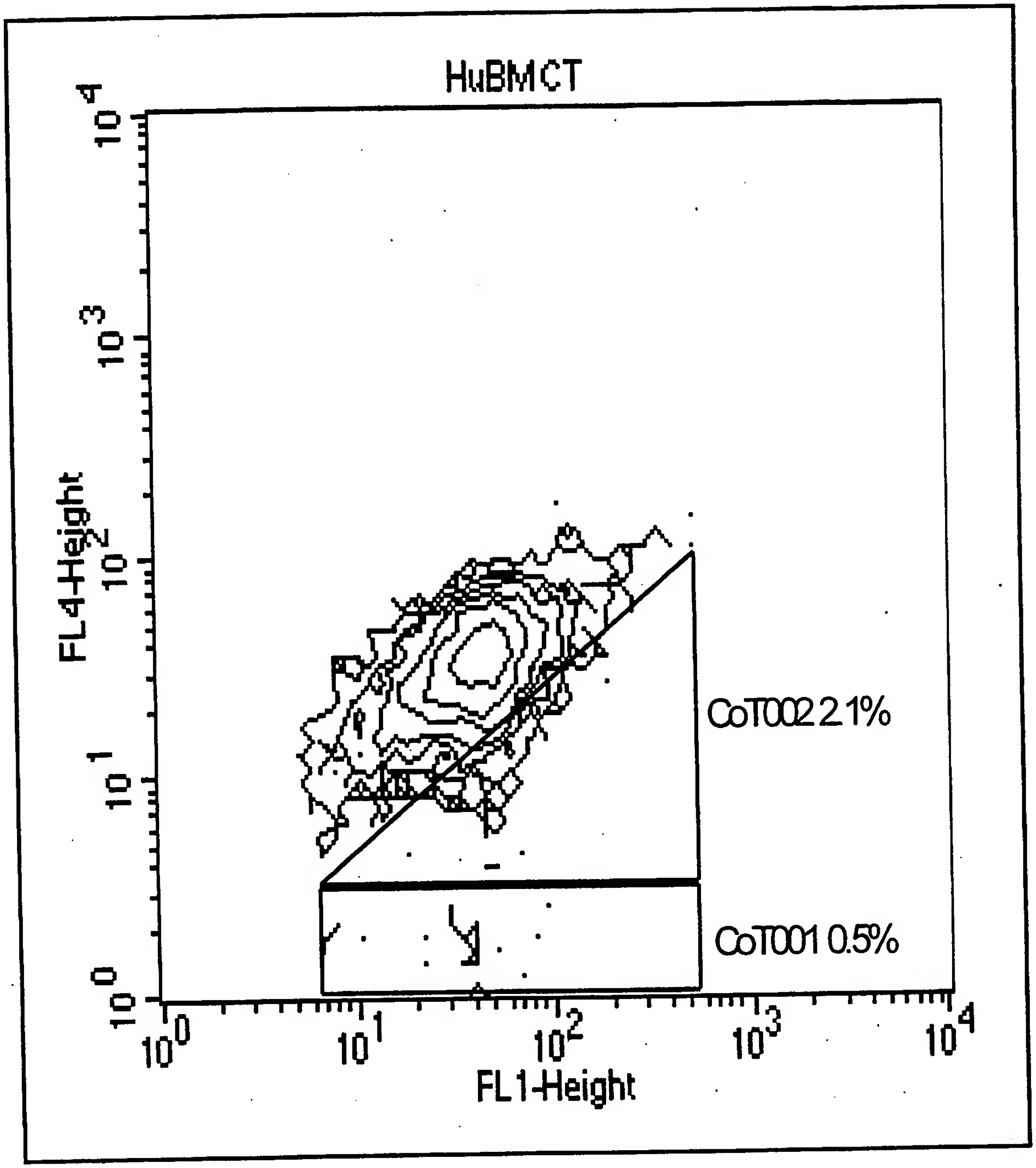
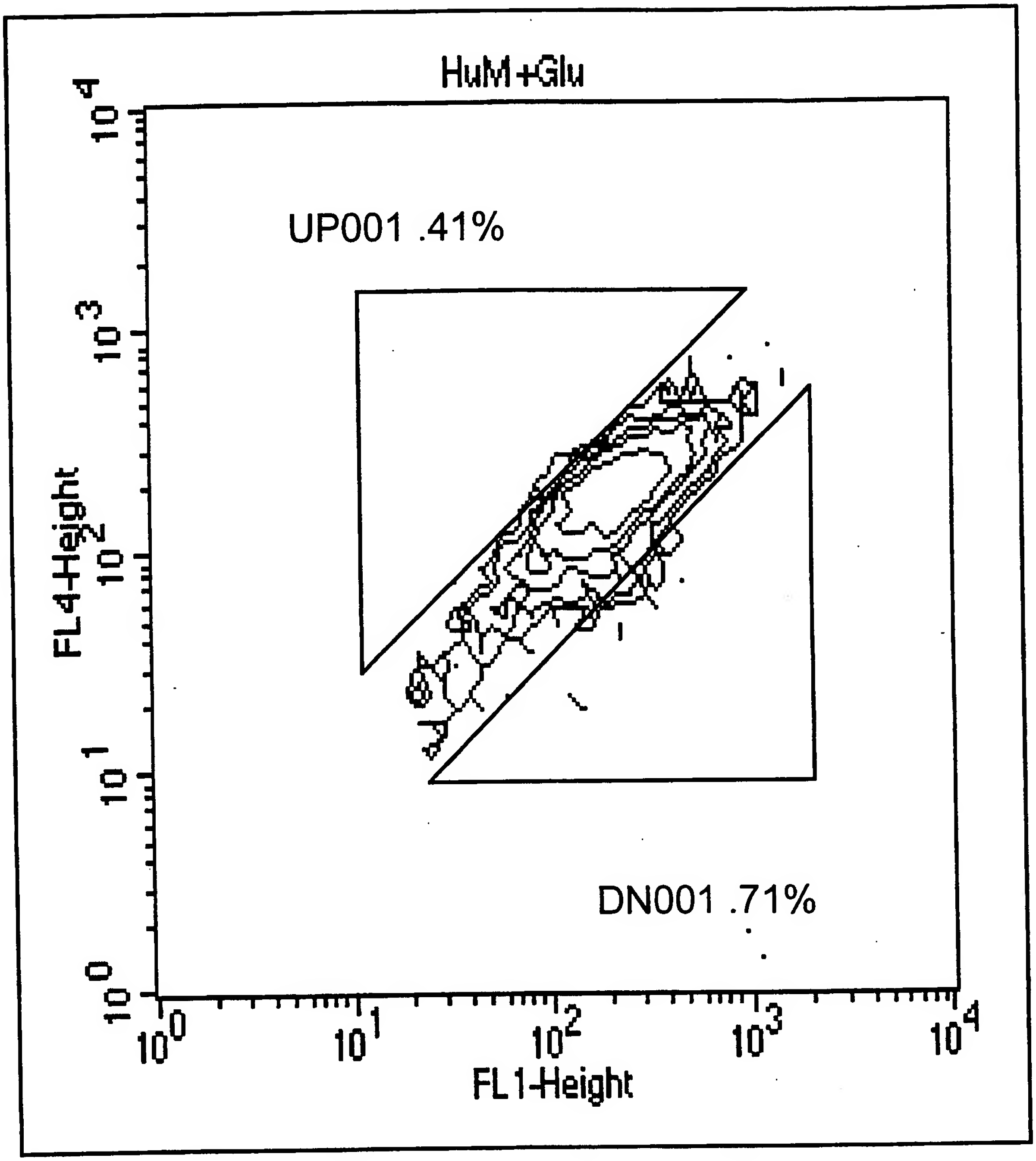

HuTH003
FL4-Height
FL1-Height
CoT003F 2.1%
CoT003E 0.2%
10^0
10^1
10^2
10^3
10^4

Fig. 7


HuBMCT
FL4-Height
FL1-Height
CoT002 2.1%
CoT001 0.5%
10^0
10^1
10^2
10^3
10^4


HuM+Glu
UP001 .41%
DN001 .71%
FL4-Height
FL1-Height
10^0
10^1
10^2
10^3
10^4

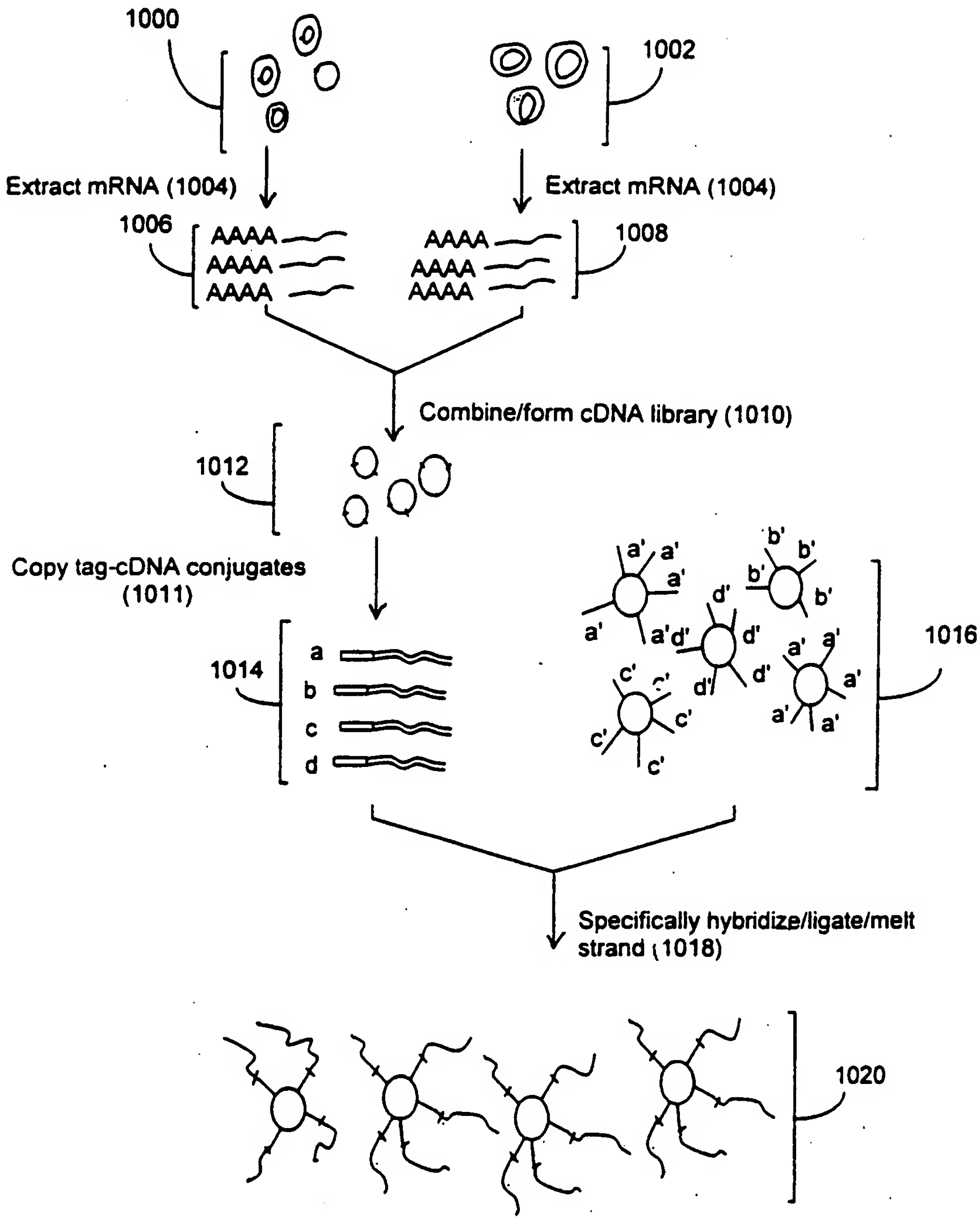

1000
1002
Extract mRNA (1004)
Extract mRNA (1004)
1006
AAAA
AAAA
AAAA
AAAA
AAAA
AAAA
1008
Combine/form cDNA library (1010)
1012
Copy tag-cDNA conjugates (1011)
1014
a
b
c
d
a'
b'
c'
d'
1016
Specifically hybridize/ligate/melt strand (1018)
1020

Fig. 10A

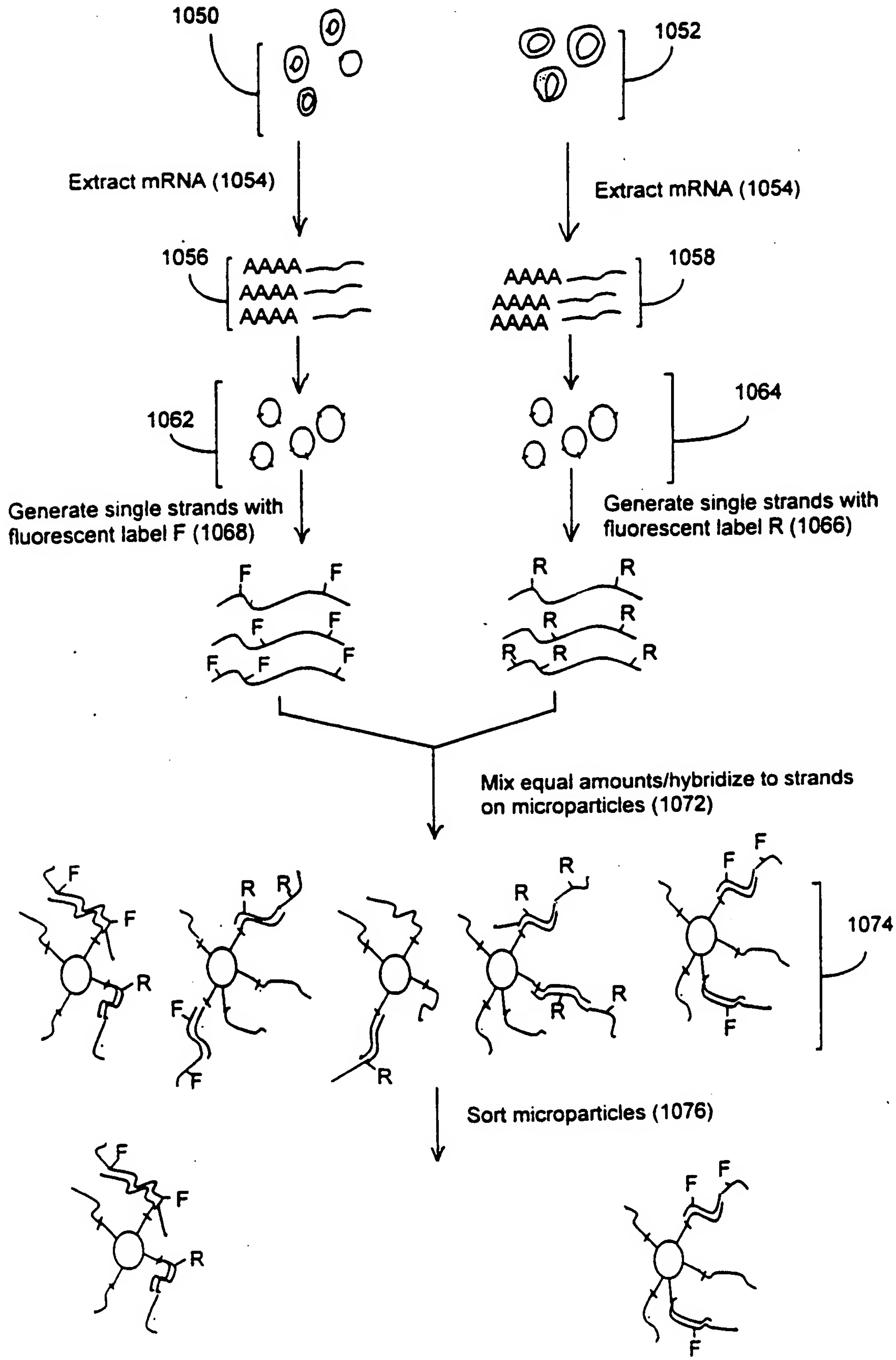

1050
1052
Extract mRNA (1054)
Extract mRNA (1054)
1056
AAAA
AAAA
AAAA
1058
AAAA
AAAA
AAAA
1062
1064
Generate single strands with fluorescent label F (1068)
Generate single strands with fluorescent label R (1066)
F
R
Mix equal amounts/hybridize to strands on microparticles (1072)
1074
Sort microparticles (1076)

Fig. 10B

**Sequence Listing

<110> Albrecht, Glenn Brenner, Sydney DuBridge, Robert B.

<120> Solid phase selection of differentially expressed genes

<130> 822-02

<140> US 09/130,546

<141> 1998-08-06

<150> US 09/005,222 <151> 1998-01-09

<160> 25

<170> Microsoft Word 5.1

<210> 1

<211> 89

<212> DNA

<213> Artificial Sequence

<220>

<221>

<222>

<223>

<400> 1

agaattcggg ccttaattaa dddddddddd dddddddddd dddddddddd 50

ddgggcccgc ataagtcttc nnnnnnggat ccgagtgat 89

<210> 2

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<221>

<222>

<223>

<400> 2

gacatgctgc attgagacga ttcttttttt tttttttttt v 41

<210> 3

<211> 52

<212> DNA

<213> Artificial Sequence

<220>

<221>

<222>

<223>

<400> 3

gacatgctgc attgagacga ttcttttttt tttttttttt vnnnngatcn 50 nn 52

<210> 4

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<221>

<222>

<223>

<400> 4

gcattgagac gattcttttt tttttttttt ttvnnnn 37

<210> 5

<211> 73

<212> DNA

<213> Artificial Sequence

<220>

<221>

<222>

<223>

<400> 5

ttaattaagg addddddddd dddddddddd dddddddddd dddgggcccg 50

cataagtctt cnnnnnngga tcc 73

<210> 6

<211> 63

<212> DNA

<213> Artificial Sequence

<220>

<221>

<222>

<223>

<400> 6

ccchhhhhhh hhhhhhhhhh hhhhhhhhhh hhhhhtcctt aattaactgg 50

tctcactgtc gca 63

<210> 7

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<221>

<222>

<223>

<400> 7

gatcacgagc tgccagtc 18

<210> 8

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<221>

<222>

<223>

<400> 8

agtgaattcg ggccttaatt aa 22

<210> 9

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<221>

<222>

<223>

<400> 9

ctacccgcgg ccgcggtcga ctctagagga tc 32

<210> 10

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<221>

<222>

<223>

<400> 10

annntacagc tgcatccctt ggcgctgagg 30

<210> 11

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<221>

<222>

<223>

<400> 11
nanntacagc tgcatccctg ggcctgtaag 30

<210> 12
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<221>
<222>
<223>
<400> 12
cnnntacagc tgcatccctt gacgggtctc 30

<210> 13
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<221>
<222>
<223>
<400> 13
ncnntacagc tgcatccctg cccgcacagt 30

<210> 14
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<221>
<222>
<223>

<400> 14
gnnntacagc tgcatccctt cgcctcggac 30

<210> 15
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<221>
<222>
<223>
<400> 15
ngnntacagc tgcatccctg atccgctagc 30

<210> 16
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<221>
<222>
<223>
<400> 16
tnnntacagc tgcatccctt ccgaacccgc 30

<210> 17
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<221>
<222>
<223>

<400> 17
ntnntacagc tgcatccctg agggggatag 30

<210> 18
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<221>
<222>
<223>
<400> 18
nnantacagc tgcatccctt cccgctacac 30

<210> 19
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<221>
<222>
<223>
<400> 19
nnnatacagc tgcatccctg actccccgag 30

<210> 20
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<221>
<222>
<223>

<400> 20
nncntacagc tgcatccctg tgttgcgcgg 30

<210> 21
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<221>
<222>
<223>
<400> 21
nnnctacagc tgcatccctc tacagcagcg 30

<210> 22
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<221>
<222>
<223>
<400> 22
nngntacagc tgcatccctg tcgcgtcgtt 30

<210> 23
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<221>
<222>
<223>

<400> 23
nnngtacagc tgcatccctc ggagcaacct 30

<210> 24
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<221>
<222>
<223>
<400> 24
nntntacagc tgcatccctg gtgaccgtag 30

<210> 25
<211> 30
<212> DNA
<213> Artificial Sequence
<220>
<221>
<222>
<223>
<400> 25
nnnttacagc tgcatccctc ccctgtcgga 30

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68

A3

(11) International Publication Number: WO 99/35293

(43) International Publication Date: 15 July 1999 (15.07.99)

(21) International Application Number: PCT/US99/00666

(22) International Filing Date: 8 January 1999 (08.01.99)

(30) Priority Data:

09/005,222	9 January 1998 (09.01.98)	US
09/130,446	6 August 1998 (06.08.98)	US

(71) Applicant (*for all designated States except US*): LYNX THERAPEUTICS, INC. [US/US]; 3832 Bay Center Place, Hayward, CA 94545 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): ALBRECHT, Glen [US/US]; 1113 Oliver Street, Redwood City, CA 94061 (US). BRENNER, Sydney [GB/GB]; 17 B St. Edwards Passage, Cambridge CB2 3PJ (GB). DUBRIDGE, Robert [US/US]; 825 Holly Road, Belmont, CA 94002 (US).

(74) Agents: ASHTON, Nina, M.; Cooley Godward LLP, Five Palo Alto Square, 3000 El Camino Real, Palo Alto, CA 94306-2155 (US) et al.

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

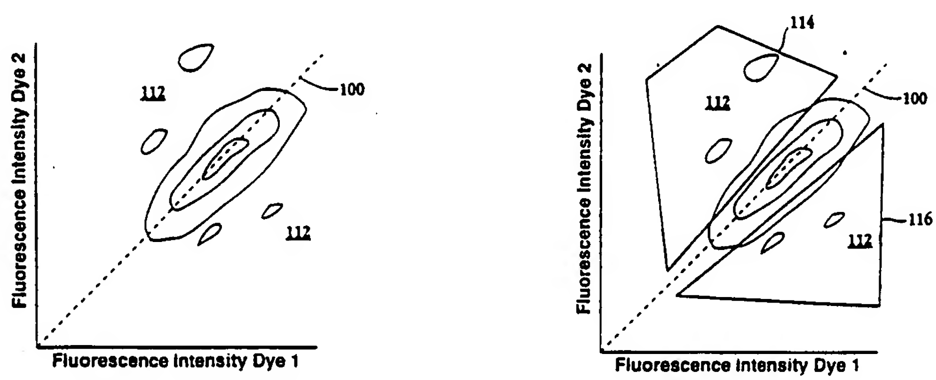
Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report: 30 September 1999 (30.09.99)

(54) Title: SOLID PHASE SELECTION OF DIFFERENTIALLY EXPRESSED GENES



(57) Abstract

The invention provides a method and materials for monitoring and isolating differentially expressed genes. In accordance with the method of the invention, differently labeled populations of DNAs from sources to be compared are competitively hybridized with reference DNA cloned on solid phase supports, *e.g.* microparticles, to provide a differential expression library which, in the preferred embodiment, may be manipulated by fluorescence–activated cell sorting (FACS). Monitoring the relative signal intensity of the different fluorescent labels on the microparticles permits quantitative analysis of expression levels relative to the reference DNA. The invention also provides a method for identifying and isolating rare genes. Populations of microparticles having relative signal intensities of interest can be isolated by FACS and the attached DNAs identified by sequencing, such as with massively parallel signature sequencing (MPSS), or with conventional DNA sequencing protocols.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATI 'AL SEARCH REPORT

Inte al Application No
PCT/US 99/00666

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SCHENA M ET AL: "QUANTITATIVE MONITORING OF GENE EXPRESSION PATTERNS WITH A COMPLEMENTARY DNA MICROARRAY" SCIENCE, vol. 270, no. 5235, 20 October 1995 (1995-10-20), pages 467-470, XP000644675 ISSN: 0036-8075 cited in the application	27-29
Y	see whole doc, esp. p.468, col.3	1-26, 30-60

Y	WO 96 12039 A (LYNX THERAPEUTICS INC) 25 April 1996 (1996-04-25) cited in the application see whole doc., esp. claims 43ff	1-26, 30-60

	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 July 1999	10/08/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Müller, F

1

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	DERISI J.L. ET AL.,: "Exploring the metabolic and genetic control of gene expression on a genomic scale" SCIENCE, vol. 278, - 24 October 1997 (1997-10-24) pages 680-686, XP002110854 the whole document ---	27-29
A	VAN NESS J ET AL: "A VERSATILE SOLID SUPPORT SYSTEM FOR OLIGODEOXYNUCLEOTIDE PROBE -BASED HYBRIDIZATION ASSAYS" NUCLEIC ACIDS RESEARCH, vol. 19, no. 12, 25 June 1991 (1991-06-25), pages 3345-3350, XP000208399 ISSN: 0305-1048 the whole document -----	

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 99/00666

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9612039 A	25-04-1996	US 5695934 A	09-12-1997
		AU 3946195 A	06-05-1996
		AU 4277896 A	06-05-1996
		CA 2202167 A	25-04-1996
		CZ 9700866 A	17-09-1997
		EP 0786014 A	30-07-1997
		EP 0793718 A	10-09-1997
		FI 971473 A	04-06-1997
		HU 77916 A	28-10-1998
		JP 10507357 T	21-07-1998
		NO 971644 A	02-06-1997
		WO 9612014 A	25-04-1996
		US 5604097 A	18-02-1997
		US 5635400 A	03-06-1997
		US 5654413 A	05-08-1997
		US 5863722 A	26-01-1999
		US 5846719 A	08-12-1998

Form PCT/ISA/210 (patent family annex) (July 1992)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 January 2001 (25.01.2001)

PCT

(10) International Publication Number
WO 01/05808 A2

(51) International Patent Classification[7]: C07K 1/00

(21) International Application Number: PCT/GB00/02809

(22) International Filing Date: 20 July 2000 (20.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9917027.6 20 July 1999 (20.07.1999) GB

(71) Applicant (*for all designated States except US*): **AFFIBODY TECHNOLOGY SWEDEN AB** [SE/SE]; Björnnäsvägen 21, S-113 47 Stockholm (SE).

(71) Applicant (*for GB only*): **GARDNER, Rebecca** [GB/GB]; Frank B. Dehn & Co., 179 Queen Victoria Street, London EC4V 4EL (GB).

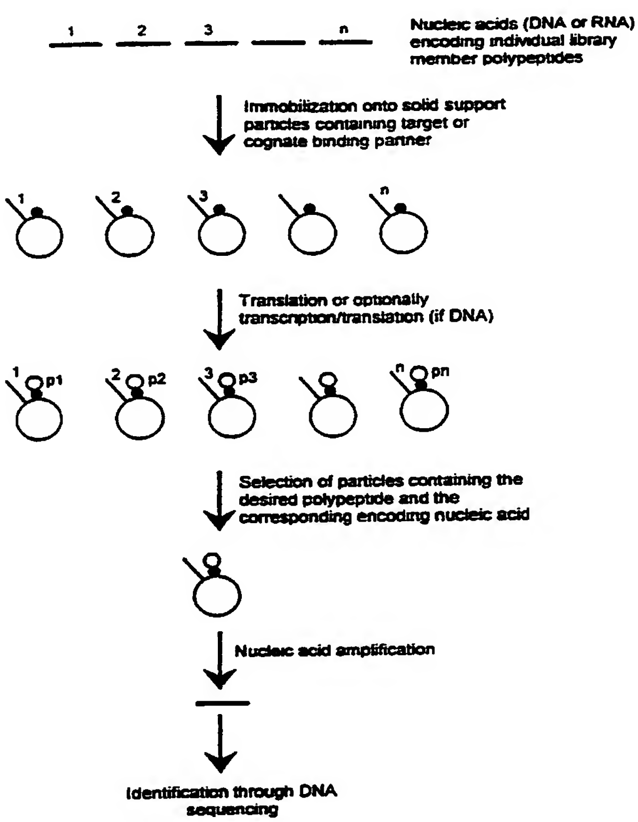
(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **NYGREN, Per-Åke** [SE/SE]; Singelgränd 36, S-178 40 Ekerö (SE). **UHLEN, Mathias** [SE/SE]; Måsvägen 8B, S-183 57 Täby (SE). **NORD, Olof** [SE/SE]; Affibody Technology Sweden AB, Björnnäsvägen 21, S-113 47 Stockholm (SE).

(74) Agents: **GARDNER, Rebecca** et al.; Frank B. Dehn & Co., 179 Queen Victoria Street, London EC4V 4EL (GB).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KR (utility model), KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

[Continued on next page]

(54) Title: *IN VITRO* SELECTION AND OPTIONAL IDENTIFICATION OF POLYPEPTIDES USING SOLID SUPPORT CARRIERS



(57) Abstract: The present invention relates to a method for the selection of one or more desired polypeptides comprising: (a) cell free expression of nucleic acid molecules immobilized on a solid support system to produce polypeptides, the solid support carrying means for biospecific interaction with at least the desired polypeptide or a molecule attached thereto; (b) separation of the solid support carrying both the desired polypeptide and the nucleic acid encoding it; and optionally (c) recovery of the said nucleic acid and/or said desired polypeptide; and molecular libraries for use in such methods.

WO 01/05808 A2

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

In vitro selection and optional identification of polypeptides using solid support carriers

This invention provides methodology for *in vitro* selection and, if desired, subsequent identification of proteins or peptides with desired properties from pools of protein or peptide variants (libraries).

Proteins and peptides, hereinafter jointly referred to as polypeptides, with desired properties such as binding affinity to a particular target molecule, catalytic activity, chemical or enzymatic activity or immunogenic activity are of great importance in many areas of biotechnology such as drug and vaccine development, diagnostic applications and bioseparation.

Recent progress in gene technology has provided the introduction of novel principles of isolating and identifying such polypeptides from large collections of variants constructed by different methods including combinatorial principles (Clackson and Wells, *Trends Biotechnol.* ***12***, pp. 173-184 [1994]). Typically, using biosynthesis for production of the library members, large pools of genes are constructed, encoding the individual library members, allowing for later selection or enrichment of desired variants using an appropriate bait molecule or chemical condition (Smith and Petrenko, *Chem. Rev.* ***97***, pp. 391-410 [1997]). For identification of selected variants, several techniques have been described to provide a physical link between the translated protein (phenotype) and the genetic information encoding it (genotype), allowing for identification of selected library members using DNA sequencing technology.

Using phage or cell display technologies, a genotype-phenotype coupling is obtained through

incorporation of the individual library members into the coat or cell surface structures respectively of phage or cells containing the corresponding gene, which is typically inserted into phage, phagemid, plasmid or viral DNA. In the construction of such libraries, the gene pools need to be transformed into a recipient cell used for biosynthesis of the corresponding proteins. The practical limitations associated with this critical step to obtain large (complex) libraries (typically above 10^9 different members) have been a driving force for the development of alternative technologies based on *in vitro* transcription and translation of genetic information, thereby avoiding the transformation step.

Examples of such technologies are ribosomal display (Mattheakis *et al.*, *Proc. Natl. Acad. Sci. USA* ***91***, pp. 9022-9026 [1994]; Hanes *et al.*, *FEBS Letters* ***450***, pp. 105-110 [1999]) and RNA-peptide fusions using puromycin (Roberts and Szostak, *Proc. Natl. Acad. Sci. USA* ***94***, pp. 12297-12302 [1997]). In ribosomal display, a gene pool (typically polymerase chain reaction (PCR) products containing signals necessary for transcription and translation) is transcribed *in vitro* to produce a corresponding pool of mRNA used for ribosome mediated translation of proteins which typically, through the absence of translational stop signals, remain physically linked to the ribosome-mRNA complex. This allows for selection of polypeptides on the basis of the characteristics of the same and identification through DNA sequencing after conversion of the ribosome-associated mRNA into DNA by the use of reverse transcriptase. However, special precautions (temperature, buffer conditions) must be taken to ensure the stability of the ribosome-mRNA-protein complexes, limiting the conditions under which selection can be performed (Jermutus *et al.*, *Curr. Opin. Biotechnol.* ***9***, pp. 534-548 [1998]; Hanes *et al.*, *op. cit.* [1999]). In

the RNA-peptide fusion system, puromycin-tagged RNA is used during translation, resulting in covalent RNA-protein/peptide links via acceptance by the ribosome of puromycin in the nascent polypeptide chain. However, new puromycin-mRNA fusions have to be prepared for each round of selection, severely limiting the efficiency of the technology (Jermutus et al., *op. cit.* [1998]; Roberts, *Curr. Opin. Chem. Biol. 3*, pp. 268-273 [1999]).

A further system has been described by Tawfik and Griffiths (Nature Biotechnology, (1998) 16; 652-656) which is cell free but seeks to mimic the effect of cells in creating compartments to link genotype and phenotype. Micelles are formed using a water-in-oil emulsion which can then be broking by mixing with ether. However, this system is not without problems, the two phase system results in several practical limitations. In order to recover the encapsulated molecules, the two phase system must be broken which is rather laborious, requiring several washes and causing a loss of material. Furthermore, the non-water components necessary to create the two-phase system might inhibit or denature biomolecules and the encapsulation itself makes it more difficult to deliver additional reagents necessary for e.g. detection or capture of specific molecular entities.

The present invention is based on the finding that by using a solid support such as a particle system as carrier of genetic information (e.g. RNA or DNA) used for identification and having coupled thereto the corresponding *in vitro* translated polypeptide, methodology linking genotype and phenotype is established. Isolation of solid support particles carrying a desired library member or members may typically be performed using sorting technology employing e.g. fluorescent labels incorporated into a target molecule or the library polypeptide members or by magnetic isolation using magnetic particles containing

an immobilized target molecule.

Thus according to one aspect of the present invention there is provided a method for the selection of one or more desired polypeptides comprising:

(a) cell free expression of nucleic acid molecules immobilized on a solid support system to produce polypeptides, the solid support carrying means for biospecific interaction with at least the desired polypeptide or a molecule attached thereto;

(b) separation of the solid support carrying both the desired polypeptide and the nucleic acid encoding it; and optionally

(c) recovery of the said nucleic acid and/or said desired polypeptide, preferably of the nucleic acid.

The selection method of the invention can be considered also as a method of enriching the desired polypeptide from a starting library of molecules containing it. 'Enrichment' referring to increasing the relative proportion of the desired polypeptide within the sample of variant molecules. Similarly, the method can be considered one by which a nucleic acid molecule of interest, i.e. which encodes the desired polypeptide is enriched.

Step (a) is cell free. The term "cell" is used in a broad sense to include cell and preferably cell-like systems and thus preferably encompasses liposomes, micelles formed by water-in-oil emulsions, gels, glass or any other multi-phase system which creates a physical barrier between one gene expression/biospecific interaction system and another. According to a preferred aspect of the present method, no actual compartmentalisation takes place, no membrane or other separation system is required to isolate individual nucleic acid molecules from one another.

The separation step (b) may advantageously be effected by interaction of the immobilized desired polypeptide with a target (e.g. biospecific) reactant

therefor which carries means permitting separation of the resulting solid support/nucleic acid/desired polypeptide/target reactant complex. Such means may, for example, comprise a label such as a fluorescence label or a magnetic particle. In this way the complex may be separated using fluorescence-activated cell sorting (FACS) technology or magnetic separation technology.

The immobilized nucleic acids may, for example, be RNA or DNA encoding individual polypeptides such as the members of a protein library. It will be appreciated that their *in vitro* translation will be effected in combination with or following *in vitro* transcription in the case of immobilised DNA.

Suitable solid supports for use in the present invention may be any of the well known supports or matrices which are currently widely used or proposed for immobilisation, separation etc. These may take the form of particles, sheets, gels, filters, membranes, fibres, capillaries, or microtitre strips, tubes, plates or wells etc., particulate solid supports being preferred. Conveniently the support may be made of glass, silica, latex or a polymeric material.

Non-magnetic polymer beads suitable for use in the method of the invention are available from Dyno Particles AS (Lillestrøm, Norway) as well as from Qiagen, Pharmacia and Serotec. However, to aid manipulation and separation, magnetic beads are preferred. The term "magnetic" as used herein means that the support is capable of having a magnetic moment imparted to it when placed in a magnetic field, and thus is displaceable under the action of that field.

Thus, using the method of the invention, after gene expression and biospecific interaction the magnetic particles may be removed onto a suitable surface by application of a magnetic field eg. using a permanent magnet. It is usually sufficient to apply a magnet to

the side of the vessel containing the sample mixture to aggregate the particles to the wall of the vessel and to pour away the remainder of the sample.

Especially preferred are superparamagnetic particles for example the well-known magnetic particles sold by Dynal AS (Oslo, Norway) as DYNABEADS, are suited to use in the present invention.

Methods for attachment of nucleic acid molecules or proteinaceous moieties such as the cognate binding partners or target molecules discussed herein to a solid support are well known in the art and many include but are not limited to chemical coupling, e.g. involving amine, aldehyde, thiol, thioether or carboxyl grous or biospecific coupling for example taking advantage of interactions between streptavidin and biotin or analogues thereof, IgG and protein A or G, HSA and protein G, glutathione S-transferase (G-ST) and glutathione, maltose and maltose binding protein, antibody and antigen (including proteins, peptides, carbohydrates and haptens), lectins and carbohydrates, hisidines and chelating groups and nucleic acid/nucleic acid hybridization.

The expressed polypeptides may advantageously be fusion proteins containing an affinity fusion partner, the solid support carrying a cognate binding partner for said affinity fusion partner as the means for biospecific interaction. Thus the expressed fusion protein will typically comprise an affinity fusion partner portion as well as the desired polypeptide or a molecular variant of the desired polypeptide from the library of molecules which contains the desired polypeptide. In this way a library of fusions proteins is generated having a variable portion which is made up of the desired polypeptide or variants thereof from the starting library and an essentially common portion, the affinity fusion partner. As appropriate, reference is made herein to molecular libraries which may be

libraries of nucleic acid molecules or libraries of polypeptides. Likewise, a library member may refer to a polypeptide or a nucleic acid molecule.

In an alternative embodiment a target molecule capable of biospecific interaction with the desired polypeptide is immobilized on the solid support as the means for biospecific interaction. In this embodiment, a library of fusion proteins may also be generated, each fusion protein incorporating a reporter protein which may conveniently be used in the separation step (b) as well as the desired polypeptide or a molecular variant of the desired polypeptide from the library of molecules which contains the desired polypeptide. Thus again, the motif of a variable portion and an essentially common portion (here the reporter protein) is provided. Each molecule within the library of fusion proteins will thus preferably have a region which is essentially the same as the corresponding region of other molecules in the library, while the variable region of each library member will differ from all or at least most of the corresponding regions of the other library members. In general one variable region will not differ significantly from some or all of the other variable regions within the library of fusion proteins. In this way the impact of minor variations in primary amino acid sequence on e.g. binding can be investigated.

Recovery of the nucleic acid(s) encoding the desired polypeptide(s) may, for example, be effected by *in vitro* amplification, e.g. by means of PCR, reverse transcriptase PCR or rolling circle amplification.

The sequence of separated and/or amplified nucleic acid(s) may be determined, e.g. by conventional sequencing techniques, thereby permitting determination of the sequence of the desired polypeptide in order to identify it.

In a further aspect of the invention the starting pool of nucleic acids encoding individual library

members may be of considerable complexity (e.g. $\geq 10^{15}$ members) (Roberts, *op. cit.* [1999]). The number of different nucleic acid species immobilized per solid phase carrier particle may be controlled in the preparation of the particles, for example through use of different concentrations of the molecule serving as anchor (for example DNA, RNA, PNA or a protein) or through pretreatment of particles with competing material. Thus the selection of discrete particles in only a single selection procedure according to the invention may result in simultaneous selection of a significantly reduced number of library members.

Performance of repeated cycles in accordance with the invention, optionally employing solid phase support particles with successively decreasing numbers of nucleic acid anchoring sites, and optionally with simultaneous dilution of the nucleic acid material, may result in gradual convergence to a limited set of library members which may be subjected to individual analysis at a clonal level in order to identify a desired polypeptide species. Where selection technology such as FACS is employed, use of different threshold values for positive selection may permit stringent selection of solid phase carrier particles containing high numbers of the desired library member.

Alternatively, after a reduction in the number of library members by separation in accordance with the invention, the enriched pool of nucleic acid sequences may be subjected to further selections using a different selection principle, such as (but not limited to) cell display, phage display, plasmid display, ribosomal display or mRNA-peptide fusions.

Thus, the method of the invention is preferably an iterative process with enrichment of the polypeptide(s) of interest occurring as more cycles are performed. While there may be some diffusion of expressed polypeptides and binding to neighbouring beads (or

regions of solid support, particles etc.), local binding to the polypeptide's own bead (or region of solid support, particle etc.) will be preferred. Thus after several cycles significant enrichment will be achieved. Method steps (a) and (b) will thus preferably be performed more than once, typically the number of cycles will be between 1 and 100, prefeably 2 to 50, more preferably 2 to 20, e.g. 5 to 10. In this way the number of variants may be very significantly limited and the relatively small number remaining can be analysed one-by-one, e.g. by ELISA, statistical analysis of clones after sequencing or Biacore analysis.

In another aspect of the invention, the selection of a solid phase support carrier carrying multiple nucleic acid species, including the desired library member, may be used to produce useful reagents without the need for identification of the particular desired library member. Thus the method may be performed in an iterative manner but stopped when the selected sample still contains a mixed population of DNA molecules; this pool of DNA fragments can be used as a "polyclonal" material, not defined at the molecular level but still useful.

In a further embodiment of the invention two different nucleic acid libraries may be immobilized on separate solid support systems and the method may be used to select and identify interacting pairs of polypeptides. Thus, for example, one of the nucleic acid libraries may encode polypeptides such as antibodies, antibody fragments, peptides or protein domains and the other may encode cDNA encoded polypeptides.

According to a further aspect of the invention is provided a molecular library comprising a solid support system having immobilised thereon a plurality of nucleic acid molecules and associated with each of said nucleic acid molecules and also immobilised on said support system means for biospecific interaction with the

expression product of one or more of said nucleic acid molecules.

The solid support system is preferably particulate and thus each particle will conveniently carry one nucleic acid molecule from the library and means for biospecific interaction with the expression product thereof. Thus the aforementioned 'association' between nucleic acid molecules and means for biospecific interaction is achieved. As discussed in more detail above, the library of nucleic acid molecules will conveniently encode fusion proteins and the means for biospecific interaction may interact, typically bind, to either the variable or common portion of said fusion protein.

In the accompanying drawings, which serve to illustrate the invention without in any way limiting it:

Fig. 1 is a schematic description of the basic concept of the invention. A pool of nucleic acid fragments encoding individual polypeptide library members are immobilized onto particles of a solid support carrier. In a DNA-based format, fragments are immobilized whereafter a coupled transcription/ translation step is performed resulting in the production of the corresponding gene products. In an RNA-based format, RNA molecules are transcriptionally produced from the DNA fragments, after which they are immobilized onto the solid support carrier, followed by a translation step resulting in the corresponding gene products. Typically, but not exclusively, the gene products are fusion proteins between polypeptide library members and an affinity fusion partner for which a cognate binding partner is present on the solid support carrier particles. Functional selection of a desired polypeptide results in isolation of particles carrying the corresponding genes (DNA or RNA) which are identified after nucleic amplification and DNA sequencing.

Fig 2 is a schematic description of the use of a solid support as carrier of coupled genetic and protein information (immobilized DNA/labelled target in solution version). A library of DNA constructs (typically but not exclusively PCR fragments) containing signals necessary for library member RNA transcription and protein translation is immobilized onto particles of a suitable carrier support (e.g. using biotin/streptavidin chemistry by incorporation of a biotin group into the DNA of the primer used for the PCR amplification and the use of streptavidin coated beads). The genetic constructs encode individual library members as genetically fused to a common affinity fusion partner (AFP) for which the cognate binding partner (CBP) is immobilized onto the particles (e.g. via suitable coupling chemistry such as streptavidin/biotin chemistry). After addition of components for *in vitro* transcription and translation (e.g. an *Escherichia coli* S30 extract), RNA (mRNA) molecules are produced which encode for the different subsequently translated protein library members. Through interaction between the immobilized binding partner and the newly translated affinity fusion partner, the individual library members are physically linked to the solid support carrier particles containing the genetic information (DNA) encoding them.

After washing, the solid support carrier particles are incubated with labelled target molecules, e.g. comprising fluoroscein isothiocyanate (FITC), allowing physical isolation of fluorescent-positive particles, for example by FACS or by magnetic separation. Thus, particles carrying complexes between the labelled target and the particle-associated library member gene product and its genetic information (DNA) are isolated.

Using e.g. PCR, the DNA fragments coupled to individual or multiple isolated particles or beads are re-amplified and used for identification of the selected

polypeptide(s) or optionally consecutive rounds of particle immobilization, *in vitro* transcription and translation followed by selection, e.g. by FACS.

Fig. 3 is a schematic representation of the use of a solid support as carrier of coupled genetic and protein information (immobilized mRNA/labelled target in solution version). From a library of genetic constructs containing signals necessary for library member transcription and protein translation, RNA (mRNA) is produced (transcription) *in vitro* and immobilized onto particles of a suitable carrier support (e.g. via hybridization between complementary sequences present in the mRNA and immobilized DNA, PNA or RNA fragments). The immobilized mRNA molecules encode individual library members as genetically fused to a common affinity fusion partner (AFP) for which the cognate binding partner (CBP) is immobilized onto the particles (e.g. via streptavidin/biotin chemistry). After addition of components for *in vitro* translation (e.g. an *Escherichia coli* S30 extract), the mRNA molecules are translated to produce the different protein library members. Through interaction between the immobilized binding partner and the newly translated affinity fusion partner, the individual library members are physically linked to the solid support carrier particles containing the genetic information (mRNA) encoding them.

After washing, the solid support carrier particles are incubated with labelled target molecules, e.g. comprising FITC, allowing physical isolation of fluorescent-positive particles, e.g. by FACS. Thus, individual or multiple particles carrying complexes between the labelled target and the particle-associated library member gene product and its genetic information (mRNA) are isolated.

Using e.g. reverse transcriptase PCR, the bead/particle-associated mRNA molecules are converted into the corresponding DNA fragments which are PCR

amplified and used for identification of the selected polypeptide(s) or optionally consecutive rounds of *in vitro* transcription, particle immobilization, *in vitro* translation followed by selection, e.g. by FACS or magnetic selection.

Fig. 4 is a schematic representation of the use of a solid support as carrier of coupled genetic and protein information (immobilized DNA/labelled binder version). A library of DNA construct (typically but not exclusively PCR fragments) containing signals necessary for library member RNA transcription and protein translation is immobilized onto discrete particles of a suitable carrier support (e.g. using biotin/streptavidin chemistry by incorporation of a biotin group into the DNA of the primer used for the PCR amplification and the use of streptavidin coated beads). The particles also carry the target molecule with which interacting library members are desired to interact. This immobilization can be achieved using e.g. standard coupling chemistries such as EDC/NHS chemistry or biotin/streptavidin chemistry. The genetic constructs encode individual library members as genetically fused to a reporter fusion partner (RFP) such as an enzyme or autofluorescent protein such as green fluorescent protein (GFP). After addition of components for *in vitro* transcription and translation (e.g. an *Escherichia coli* S30 extract), RNA (mRNA) molecules are produced which encode for the subsequently translated different protein library members. Through interaction between the immobilized target molecule and the newly translated library member, individual library members capable of interaction with the solid support immobilized target molecule are physically linked to the solid support carrier particles containing the genetic information (DNA) encoding them.

After washing, the solid support carrier particles are sorted, e.g. using FACS technology or magnetic

separation, to isolate individual or multiple particles carrying complexes between the immobilized labelled target and the particle-associated library member gene product. Thus, particles carrying complexes between the labelled target and the particle-associated library member gene product and its genetic information (DNA) are isolated.

Using PCR, the DNA fragments coupled to discrete isolated beads are re-amplified and used for identification of the selected polypeptide(s) or optionally consecutive rounds of particle immobilization, *in vitro* transcription and translation followed by separation, e.g. by FACS or magnetic selection.

Fig. 5 is a schematic representation of the use of a solid support as carrier of coupled genetic and protein information (immobilized mRNA/labelled binder version). From a library of genetic constructs containing signals necessary for library member transcription and protein translation, RNA (mRNA) is produced (transcription) *in vitro* and immobilized onto particles of a suitable carrier support (e.g. via hybridization between complementary sequences present in the mRNA and immobilized DNA, PNA or RNA fragments). The particles also carry the target molecule with which library members are desired to interact. This immobilization may be obtained using e.g. standard coupling chemistries such as EDC/NHS chemistry or biotin/streptavidin chemistry. The genetic constructs (mRNA) encode individual library members as genetically fused to a reporter fusion partner (RFP) such as an enzyme or autofluorescent protein such as green fluorescent protein (GFP). After addition of components for *in vitro* translation (e.g. an *Escherichia coli* S30 extract), mRNA molecules are translated to produce the different protein library members. Through interaction between the immobilized target molecule and the newly

translated library member, individual library members capable of interaction with the solid support immobilized target molecule are physically linked to the solid support carrier containing the genetic information (mRNA) encoding them.

After washing, the solid support carriers are sorted, e.g. using FACS technology, to isolate individual or multiple particles carrying complexes between the immobilized labelled target and the particle-associated library member gene product. Thus, particles carrying complexes between the labelled target and the particle-associated library member gene product and its genetic information (mRNA) are isolated.

Using e.g. reverse transcriptase PCR, the bead/particle-associated mRNA molecules are converted into the corresponding DNA fragments which are PCR amplified and used for consecutive rounds of *in vitro* transcription, particle immobilization, *in vitro* translation followed by selection, e.g. by FACS.

Fig. 6 illustrates the experimental set-up for Example 1. Paramagnetic particles coated with streptavidin were firstly incubated with biotinylated human serum albumin (HSA), resulting in robust anchoring of HSA. Separate aliquots were subsequently incubated with either (A) protein ABD-Z, a genetic fusion protein between a serum albumin binding protein (ABD) derived from streptococcal protein G and an immunoglobulin binding protein (Z) derived from staphylococcal protein A, followed by incubation with fluorescent isothiocyanate (FITC) conjugated polyclonal goat IgG antibodies, or (B) with the FITC conjugated goat IgG antibodies directly.

Fig. 7 is a photograph from UV-microscopy analyses of streptavidin-coated beads/particles containing streptavidin/biotin chemistry-immobilized biotinylated human serum albumin. (A) Particles incubated with FITC-conjugated polyclonal goat IgG antibodies after having

first been subjected to a solution containing the fusion protein Z-ABD. (B) Particles incubated with FITC-conjugated polyclonal goat IgG antibodies only.

Fig. 8 is a schematic representation of the use of the invention for selecting interacting polypeptide pairs through the crossing of two different libraries. Two pools of nucleic acid fragments encoding different polypeptide libraries are separately immobilized onto particles of solid support carrier systems. In a DNA-based format, fragments are immobilized whereafter a coupled transcription/translation step is performed resulting in the production of the corresponding gene products. In an RNA-based format, RNA molecules are transcriptionally produced from the DNA fragments, after which they are immobilized onto the solid support carrier, followed by a translation step resulting in the corresponding gene products. Typically, but not exclusively, the gene products are fusion proteins between polypeptide library members and an affinity fusion partner for which a cognate binding partner is present on the particles. The different libraries are differently labelled, e.g. using two fluorophores having different excitation spectra. Biospecific interactions between members of the different polypeptide libraries are detected as double-labelled particle pairs. For identification, the nucleic acids present on the isolated particles encoding the corresponding genes are analyzed by DNA sequencing.

Fig. 9 is a schematic description of the construction of the plasmids pGEM-SD-K-FLAG-Z_{wt} and pGEM-SD-K-FLAG-Z_{IgA}, designed for use as template for the amplification of PCR products for cell free transcription and translation of either free or bead-immobilized DNA/RNA.

Fig. 10 is a radiograph obtained after SDS-PAGE analysis under reducing conditions of proteins synthesized using a cell free extract supplemented with

[35S]methionine and PCR products produced with primers NOOL-12 and NOOL-13 using different plasmids as templates. Lane 1: pGEM-SD-K-FLAG-Z_{wt}, lane 2: pGEM-SD-K-FLAG-Z_{IgA}. A marker with 14C-labeled proteins was used as size reference (prod. no. CFA756, Amersham Pharmacia Biotech, Uppsala, Sweden). Arrows indicate the positions of reference proteins with molecular weights of 14.3, 20.1 and 30.0 kDa, respectively.

Fig. 11 is an overlay plot from a comparative FACS analysis of anti-FLAG BioM5 antibody-coated beads subjected to a FLAG-Z_{wt} PCR product transcription/ translation mixture and negative control beads treated in the same way but not coated with anti-FLAG BioM5 antibodies.

Fig. 12 is an overlay plot from a comparative FACS analysis of anti-FLAG BioM5 antibody and PCR product doubly coated beads, subjected to a transcription/ translation mixture, followed by detection. The picture shows the analysis of two different sets of beads containing either FLAG-Z_{wt} or FLAG-Z_{IgA} encoding PCR products subjected to the analysis.

Fig. 13 (A) is a schematic representation of the presence of a Mlu I restriction site in the PCR product obtained by PCR amplification using primers NOOL-12 and NOOL-13 on a pGEM-SD-K-FLAG-Z_{wt} plasmid template. In contrast, no Mlu I site is present in the PCR product obtained by PCR amplification using primers NOOL-12 and NOOL-13 on a pGEM-SD-K-FLAG-Z_{IgA} plasmid template. Also shown are the sizes of the cleavage products obtained after incubation of the FLAG-Z_{wt} fusion protein encoding PCR product after incubation with Mlu I.

(B) are photographs showing agarose gel electrophoresis analyses of PCR products obtained by PCR amplification of different samples taken before or after FACS-based enrichments. Lane 1: Beads containing FLAG-Z_{wt} encoding PCR product only; lane 2: Beads containing FLAG-Z_{wt} encoding PCR product only. Resulting

PCR product subjected to incubation with Mlu I; lane 3: Beads containing FLAG-Z_{IgA} encoding PCR product only; lane 4: Beads containing FLAG-Z_{IgA} encoding PCR product only. Resulting PCR product subjected to incubation with Mlu I; lane 5: Beads containing a 1:1 mixture of FLAG-Z_{wt} and FLAG-Z_{IgA} encoding PCR products. Sample from before FACS enrichment experiment; lane 6: Beads containing a 1:1 mixture of FLAG-Z_{wt} and FLAG-Z_{IgA} encoding PCR products. Resulting PCR product subjected to incubation with Mlu I. Sample from before FACS enrichment experiment; lane 7: Sample from beads sorted in FACS enrichment experiment; lane 8: Sample from beads sorted in FACS enrichment experiment. Resulting PCR product subjected to incubation with Mlu I. Flanking lanes with size markers (phage l DNA cleaved with Pst I, Amersham Pharmacia Biotech, Uppsala, Sweden) are labeled M.

Fig. 14 Top: is an overlay plot of intensity recordings of tracks corresponding to lanes 6 (dashed line) and 8 (solid line) in figure 13. The relative intensity is shown as a function of the migration coordinate. Bottom: shows digitally excised tracks from the gel image corresponding to lane 6 and 8 from the gel shown in figure 13. A relative shift of intensity towards the smaller molecular weight cleavage products is observed for the sample obtained by PCR amplification of nucleic acids present on beads collected in the FACS enrichment (track corresponding to lane 8).

In a representative embodiment of the method of the invention a pool of gene fragments (Figs. 2-5) containing the DNA encoding different polypeptide library members is prepared using standard DNA technology, for example as described by Nord *et al.*, *Prot. Engineering* ***8***, pp. 601-608 [1995] and Nord *et al.*, *Nature Biotechnol.* ***15***, pp. 772-777 [1997]. The gene fragments should include a first sequence corresponding to a suitable RNA polymerase promoter sequence, such as

E. coli phage T7 promoter, T3 promoter, SP6 promoter, *lac* promoter, *lac* UV5 promoter, *ara* B promoter, *trp* promoter, staphylococcal protein A promoter, or viral promoters such as Raus Sarcoma Virus (RSV) promoter, and Cytomegalo virus (CMV) late and early promoters to function as signals for transcription of the DNA fragment into mRNA using a suitable extract such as an S30 extract of *E. coli* for promoters of *E. coli* or prokaryotic origin or a reticulocyte extract or wheat germ extract for promoters of eukaryotic origin (coupled systems) or by a first transcriptional step using a preparation of purified suitable RNA polymerase, separated from a later translational step (uncoupled system) in which the mRNA templates are used for translation of the genetic information into the corresponding polypeptides.

In one aspect of the invention, the promoter sequence is followed by a sequence encoding an affinity fusion partner (AFP), employed for binding a cognate binding partner immobilized onto a solid phase carrier particle. This affinity fusion partner may for example be the albumin binding region of streptococcal protein G or derivatives thereof, the immunoglobulin binding protein A or derivatives thereof, maltose binding protein, glutathione S-transferase, FLAG peptide, Bio-tag (biotinylated peptide), hexahistidyl sequence, c-myc tag, or any other polypeptide for which a suitable cognate binding partner is available. The gene fragments should each also contain the gene encoding an individual library member polypeptide, in translational frame with the affinity fusion partner polypeptide if used. Alternatively, the gene encoding the affinity fusion partner may be positioned after the gene for the polypeptide library member.

In one aspect of the invention, the sequence encoding the individual library member polypeptide is either preceded or followed by a sequence encoding a

suitable reporter polypeptide, such as green fluorescent protein (GFP), alkaline phosphatase, luciferase, horse radish peroxidase (HRP) or β-galactosidase.

In one aspect of the invention, the gene fragments contain a suitable chemical group (e.g. biotin or digoxin) introduced e.g. by PCR amplification using a primer or nucleotides labelled with the group. This group is used for anchoring the DNA fragment onto solid support particles coated with a suitable cognate binding partner, such as streptavidin or anti-digoxin antibody(ies) (Figs. 2 and 4).

In another aspect of the invention, a pool of transcribed mRNA is immobilized onto the solid support particles via a suitable attachment moiety. This moiety may for example be a nucleotide sequence at the 5'- or 3'- end of the mRNA, for which a complementary sequence of RNA, DNA or PNA is immobilized onto the solid support particles (Figs. 3 and 5).

After immobilization of DNA fragments onto the solid support particles, a transcription step is performed using a suitable RNA polymerase depending on the promoter used for the construction of the fragments. The thereby transcribed mRNA is employed for translation of the genetic information into the corresponding polypeptides which are bound to the solid support particles by biospecific interaction with either an immobilized cognate binding partner for an affinity fusion partner encoded in translational frame with the polypeptide or via recognition of a target molecule immobilized onto the particle. For the translation a suitable extract or pure components may be used such as an E. coli S30 extract, a rabbit reticulocyte extract or a reconstituted mixture of purified essential components of a translation machinery. Suitable particles may for example be made of polystyrene or any other polymer or mixtures of polymers, cellulose, hydroxyapatite, sepharose, dextran or silica.

After immobilization of mRNA molecules onto solid support particles, the translation of these into the corresponding proteins is performed as described above. The thereby produced polypeptides are bound to the solid support particles by biospecific interaction with either an immobilized cognate binding partner for an affinity fusion partner encoded in translational frame with the polypeptide or via recognition of a target molecule immobilized onto the particle.

To circumvent cross-over reactions, i.e. the binding of a translated polypeptide fusion protein molecule to a cognate binding partner or target molecule present on a solid support particle not also carrying the genetic information (DNA or RNA) encoding the polypeptide, the mixture may be diluted so as to prevent close proximity between particles.

Selection of particles containing a desired polypeptide or group of polypeptides may be performed by direct isolation, for example in an FACS scanner if the target is labelled with a fluorophore or if the polypeptide is genetically fused to a fluorescent protein such as green fluorescent protein. A different selection method is to use magnetic principles, using magnetic (or paramagnetic) particles coated with the target molecule of interest (Figs. 1 and 2). Alternatively, particles labelled via a specific interaction between a library member polypeptide gene product may be physically isolated using e.g. a UV-microscope.

Selection may be performed on the basis of functional properties of the encoded polypeptides, such as binding to a desired target (antibodies or other proteins or peptides, carbohydrates, organic molecules, cells, viruses, plants etc.), catalytic activity, or through proteolytic or chemical stability under certain chemical conditions.

After isolation of particles carrying a polypeptide

with the desired characteristics, the nucleic acid information (DNA or RNA) present on the same particles is amplified (if necessary) by *in vitro* nucleic acid amplification methods such as reverse transcriptase PCR (if RNA), PCR (if DNA), or rolling circle replication.

If necessary, the procedure may be repeated for additional cycles of direct DNA immobilization or RNA immobilization after *in vitro* transcription of re-amplified particle-bound nucleic acids. If further variation is desired for the next round of selection, the amplification conditions or polymerase(s) may be chosen to introduce mutations into the next pool of DNA fragments.

In yet another aspect of the invention two different libraries of polypeptides are investigated for interacting pairs (Fig. 8). Particles corresponding to a library of e.g. cDNA encoded polypeptides are mixed with particles carrying members of a polypeptide library of, for example, cDNA encoded proteins, antibodies or fragments thereof, peptides or protein domains. The particles used for the immobilization of the nucleic acids are prepared such that they contain two different labels, one for each library. Isolation of interacting pairs of polypeptides resulting from biospecific interactions are isolated by e.g. FACS technology, employing detection of double-labelled particle pairs.

The method of the invention has several advantages over existing selection systems using an *in vivo* polypeptide biosynthesis step, since there is no need for transformation of the genetic material into a recipient cell. The only limitation with respect to library size (complexity) is the binding capacity of the solid support system. Furthermore, the present *in vitro* selection system uses a robust solid support as the linkage between genotype and phenotype, enabling harsh conditions to be used when selecting ligands with high affinity towards a given target molecule. As a

consequence of the nucleic acids being directly immobilized on the solid support they may easily be recovered; thus, for example, if the solid support comprises magnetic beads these may be removed from the transcription/translation mixture with a magnet, thus lowering the risk of contamination with non-immobilized nucleic acids.

The following non-limitative Example serves to illustrate the invention.

STANDARD PROCEDURES:

Cloning and PCR amplifications:

Standard cloning work including plasmid preparations, restriction enzyme cleavage and ligations etc. was performed as described in (Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular cloning: a laboratory manual, 2nd edn., Cold Spring Harbor Laboratory, New York, 1989) and according to suppliers recommendations. Restriction enzymes and ligase were purchased from either MBI Fermentas, Vilnius, Lithuania or New England Biolabs, MA, USA) PCR amplifications using plasmids or bead-immobilized PCR products as templates were performed in a GeneAmp® PCR system 9700 (PE Biosystems, Foster City, CA, USA), using standard conditions. As primers, oligonucleotides from Table 1 were used as specified in the examples. Typically, 5 pmoles of primers were used in a 30-cycle PCR amplification using a buffer consisting of 0.2 mM deoxyribonucleoside triphosphates (dNTPs), 50 mM KCl, 2 mM MgCl2, 10 mM Tris-HCl (pH 8.5), 0.1% Tween 20 and 0.1 units of AmpliTaq® DNA polymerase (PE Biosystems). A standard PCR cycle had the follwing settings: 15 s 94°C, 20 s 55°C, 1 min 72°C. Standard agarose gel electrophoresis analyses of nucleic acids were performed using ethidium bromide for staining. E. coli cells used for cloning and plasmid preparations were RR1DM15 (Rüther, U. Nucl.

Acids Res. 10: 5765-5772, 1982).

Table 1. List of oligonucleotide primers.

Name	Sequence 5'-3'
NOOL-6	GGGGGGAAGCTTGGGGGGGCCATGGCTTTAGCTGAAGCTAAAGTCTTAG
NOOL-7	CTTTGTTGAATTTGTTGTCTACGCTCGAGCTAGGTAATGCAGCTAAAATTTCAT
NOOL-8	ATGAAATTTTAGCTGCATTACCTAGCTCGAGCGTAGACAACAAATTCAACAAAG
NOOL-9	GGGGGAATTCTTATTATTTCGGCGCCTGAGCATCAT
NOOL-10	GGGGGGAAGCTTGGGGG
NOOL-11	GGGGGAATTCTTATTATTTCG
NOOL-12	GTTGTGTGGAATTGTGAG
NOOL-13	Biotin-AAGTTGGGTAACGCCAGG
SD KOZAK-1	AGCTTAATAATTTTGTTTAACTTTAAGAAGGAGATATAGC
SD KOZAK-2	CATGGCTATATCTCCTTCTTAAAGTTAAACAAAATTATTA
FLAG-1	CATGGACTACAAAGATGACGATGATAAAAGC
FLAG-2	TCGAGCTTTTATCATCGTCATCTTTGTAGTC

Recombinant protein production:

E. coli cells used for expression were either RR1DM15 (Rüther, U. Nucl. Acids Res. 10: 5765-5772, 1982) or BL21DE3 (Novagen, Madison, WI, USA). Osmotic shock procedures were performed as described earlier (Nygren et al., J. Mol. Recognit. 1:69-74, 1988). Affinity chromatography purifications of proteins on HSA and IgG-Sepharose resins were performed as described earlier (Nygren et al., J. Mol. Recognit. 1:69-74, 1988). Human polyclonal IgG was supplied by Pharmacia and Upjohn AB, Stockholm.

Protein biotinylation:

Human serum albumin (HSA) (prod no. A-8763, Sigma) was biotinylated using EZ-LinkTM Sulfo-NHS-LC-Biotin kit (prod no. 21335, Pierce Chemical Company, Rodeford, IL, USA).

Cell free transcription and translation of PCR fragments:

PCR products as indicated were subjected to cell free transcription and translation using a commercial E. coli S30 extract system for linear DNA (prod no. L1030, Promega, Madison, WI, USA) according to the instructions by the manufacturer. For coupled transcription/ translation of free (non-immobilized) PCR products, typically, 10-70 ng of PCR product was mixed with 50 μl of cell extract and incubated for 1 h at 25°C. In other experiments, PCR products were immobilized onto streptavidin coated microbeads (M280-SA, Dynal, Norway or Bang Laboratories, prod. no. CP01N/004109, where indicated). Such beads had previously been incubated with a 1.89 mg/ml solution of biotinylated BioM5 antibody (prod no. F-2922, Sigma, Saint Louis, MO, USA) directed to a FLAG peptide for affinity capture of FLAG peptide-tagged proteins. Typically, 10 ng of PCR product were mixed with 1 mg of BioM5-containing beads, which were subsequently washed two times before a coupled transcription/translation reaction was performed using 25 μl of E. coli extract.

Protein gel electrophoresis:

Sodium dodecylsulphate polyacrylamide gel electrophoresis of proteins (SDS-PAGE) under reducing conditions was performed using the Phast system (Amersham Pharmacia Biotech, Uppsala, Sweden) or in a Novex Xcell II (San Diego, CA, USA), as described by the respective suppliers.

DNA sequencing:

DNA sequencing was performed by cycle sequencing (Carothers et al., BioTechniques 7:494-499, 1989; Savolainen, P., et al., Mol. Biol. Evol. 17:474-488, 2000) using ThermoSequenase DNA polymerase (Amersham Pharmacia Biotech) and primers as indicated. Sequencing reactions were loaded onto a ABI Prism 377XL instrument (PE Biosystems, Foster City, CA, USA).

Fluorescence-activated cell sorting (FACS) experiments: FACS analyses were performed with either a FACSCalibur, FACScan or a FACSVantage SE instrument (Becton Dickinson, Oxnard, USA).

Where indicated, horseradish peroxidase-conjugated antibodies were used for signal amplifications, using a fluorescein tyramide reagent (Boehringer Mannheim, Germany) as described by Anton and coworkers (Anton et al., J. Histochem. Cytochem. 46:771-777, 1998).

Example 1

Discrimination between solid support particles labelled with fluorescent proteins through a biospecific interaction and control solid support particles

Approximately 2 mg of streptavidin coated particles (M280-SA, Dynal, Norway) were incubated with 30 μl of a 2 mg/ml solution in PBS buffer (0.15 M NaCl, 20 mM phosphate, pH 7.2) of human serum albumin (HSA) (Sigma art. No. A-8763) biotinylated using a protein biotinylation kit (Pierce art. No. 21335) according to the manufacturers instructions. Particles were then either directly incubated with polyclonal goat IgG antibodies, labelled with FITC (Sigma art. No. F-9887) or first incubated with 30 μl of a 2 mg/ml solution in PBS of a fusion protein (Z-ABD) between a serum albumin binding protein (ABD) derived from streptococcal protein G and a immunoglobulin binding protein (Z) derived from staphylococcal protein A produced and HSA-affinity purified as previously described (Nord *et al., op. cit.* [1995], and [1997]). Between each incubation multiple (5-10) washings with PBS were performed to remove non-specifically bound proteins.

To investigate whether discrimination was possible between particles labelled by the FITC-labelled goat

polyclonal antibodies via a biospecific interaction to the Z moiety of the Z-ABD fusion protein and particles not incubated with the Z-ABD fusion protein and thus incapable of binding the goat antibody, particles were analysed by UV-microscopy using a Olympus BH2-RFCA microscopy at an excitation wavelength of 495 nm. The results shown in Fig. 5 show that a clear difference in fluorescent intensity can be seen between the two differently treated pools of particles (Fig. 5A and 5B). This shows that the result of a biospecific interaction between an (ABD-HSA)-immobilized fusion protein and a labelled target protein added in solution can be observed.

Example 2

Assembly and cloning of genetic constructs for cell free transcription and translation experiments

To be able to obtain PCR products encoding relevant proteins or protein library members and suitable or cell free transcription and translation experiments using solid supports as carriers for both nucleic acids and their corresponding encoded proteins, a genetic construct was assembled in the plasmid vector pGEM-4Z (Figure 9). In a splice overlap extension (SOE) PCR reaction using primers NOOL-10 and NOOL-11 (table 1), two gene fragments encoding an albumin binding protein (APB) (Larsson, et al., Prot. Expr. Purif. 7:447-457, 1996) and the Z domain (Z_{wt}) (Nilsson et al., Prot. Engineering 1:107-113, 1987), respectively, were joined. The two fragments had previously been produced by separate PCR reactions using pT7-ABPc (ABP) (Larsson, et al., Prot. Expr. Purif. 7:447-457, 1996) (primers NOOL-6 and NOOL-7, table 1) or pKN1-Z_{wt} (Nord et al., Prot. Engineering, 8:601-608, 1995) (primers NOOL-8 and NOOL-9, table 1) as plasmid templates, respectively. In the SOE reaction, two fragments were joined resulting in an ABP-(Ser)3-Z_{wt} encoding gene fragment comprising in

the 5´-end recognition sites for the two enzymes Hin dIII and Nco I, and in the 3´-end two translational stop codons and a recognition site for the restriction enzyme Eco RI (Figure 9). This fragment was inserted by ligation as a Hin dIII-Eco RI fragment into the plasmid pGEM-4Z, cleaved with the same enzymes, resulting in the construct pGEM-ABP-Z_{wt}.

A fragment was assembled by the annealing of the two oligonucleotides SD KOZAK-1 and SD KOZAK-2 (table 1), resulting in a 40 bp fragment comprising an E. coli Shine Dalgarno (SD) sequence (for efficient E. coli translation) and a Kozak sequence (to facilitate expression in cell extracts from mammalian sources), flanked by Hin dIII and Nco I restriction sites (Figure 9). This fragment was inserted by ligation into pGEM-ABP-Z cleaved with Hin dIII and Nco I, resulting in the plasmid vector pGEM-SD-K-ABP-Z_{wt}. This vector was subsequently cleaved with enzymes Nco I and Xho I, releasing the ABP encoding fragment. The thereby obtained vector fragment was ligated to a FLAG peptide encoding gene fragment, previously obtained by annealing the two oligonucleotides FLAG-1 and FLAG-2 (table 1), resulting in the vector pGEM-SD-K-FLAG-Z. This vector thus encodes a FLAG-Z_{wt} fusion protein, linked by a (Ser)3 linker (Figure 9). The vector also contains an upstream T7 promoter which is capable of driving the transcription of the FLAG-Z_{wt} fusion protein gene by the action of T7 RNA polymerase. From this vector, any suitable gene fragment inserted between the Xho I and Eco RI sites can be transcribed as an mRNA operatively linked to a SD sequence, a Kozak sequence and a FLAG peptide encoding part. In addition, using primers NOOL-12 and NOOL-13 (table 1), PCR products can be obtained which are suitable for T7 RNA polymerase driven transcription and are biotinylated in their 3´-ends, suitable for immobilization on e.g. streptavidin coated

surfaces and other solid supports.

To construct the vector denoted pGEM-SD-K-FLAG-Z_{IgA}, in which the Z_{wt} encoding gene fragment has been substituted for a gene fragment encoding the human IgA-binding protein Z_{IgA} (Gunneriusson et al., J. Bact. 1999), a Z_{IgA} encoding gene fragment was amplified using primers NOOL-8 and NOOL-9 using a plasmid pKN1-Z_{IgA} template (Gunneriusson et al., J. Bact. 1999). The resulting PCR product was cleaved with restriction enzymes Xho I and Eco RI and inserted into the vector pGEM-SD-K-FLAG-Z_{wt}, previously cleaved with the same enzymes. The resulting vector pGEM-SD-K-FLAG-Z_{IgA} thus encodes a FLAG-Z_{IgA} fusion protein, linked by a (Ser)3 linker (Figure 9).

Example 3

Cell free transcription/translation of FLAG-Z_{wt} and FLAG-Z_{IgA} fusion proteins from their respective PCR products.

Using the plasmid vectors pGEM-SD-K-FLAG-Z_{wt} and pGEM-SD-K-FLAG-Z_{IgA}, respectively, for PCR amplifications using the primers NOOL-12 and NOOL-13 (table 1), PCR products were obtained of which approx. 70 ng were subjected to a one hour cell free transcription/ translation at 25°C using 50 µl of an E. coli S30 cell extract (L1030, Promega, MA, USA), supplemented with [^{35}S]methionine and 1600 units of T7 RNA polymerase. Samples of the different transcription/translation mixtures were analyzed by 10% NuPAGE (Novex, San Diego, CA, USA) under reducing conditions through the addition of 50 mM DTT (final concentration) in the sample loading buffer (NuPAGE LDS sample buffer, Novex) followed by exposure of the gel to a film (Kodak XOMAT-AR, 18x24 cm) at -70°C over night. The development of the film revealed radioactive protein of expected sizes (~8 kDa) for both the FLAG-Z_{wt} and the FLAG-Z_{IgA} encoding PCR products (Figure 10). This shows that the constructed

plasmid vectors pGEM-SD-K-FLAG-Z_{wt} and pGEM-SD-K-FLAG-Z_{IgA}, both were suitable for use as templates for the amplification of PCR products capable of directing a T7 RNA polymerase driven transcription of mRNA which could be used for cell free translation of FLAG-Z_{wt} and FLAG-Z_{IgA} fusion proteins in an E. coli S30 extract.

Example 4

Immobilization of FLAG-Z_{wt} and FLAG-Z_{IgA} fusion proteins on anti-FLAG antibody-containing beads

To investigate the functionality of the FLAG peptide moieties of the fusion proteins FLAG-Z_{wt} and FLAG-Z_{IgA}, reaction mixtures obtained from production of the two fusion proteins from their respective PCR products using cell free transcription/translation as described in example 3 were mixed for three hours at room temperature with streptavidin coated M-280-SA dynabeads (Dynal, Norway) (50 mg) previously incubated with 5 μl of a 1.89 mg/ml solution of biotinylated anti-FLAG BioM5 monoclonal antibodies (Sigma) in PBS (0.15 M NaCl, 20 mM phosphate, pH 7.2). In the experiment, beads which had not been incubated with the biotinylated anti-FLAG BioM5 antibody solution were also included (control). The beads were subsequently washed with PBST (PBS with 0.1% Tween 20) and analyzed using a Beckman LS6000 SC scintillator (Beckman-Coulter, Fullerton, CA, USA), under standard conditions using scintillation buffer. The measured signals from anti-FLAG BioM5-coated beads subjected to the transcription/translation mixtures corresponding to the FLAG-Z_{wt} and FLAG-Z_{IgA} fusion protein, respectively, were significantly higher compared to the negative controls (Table 2). This shows that fusion proteins, here exemplified by the two fusion proteins FLAG-Z_{wt} and FLAG-Z_{IgA}, can be produced from their respective PCR products by cell free transcription/translation containing a functional

affinity fusion partner, here exemplified by the FLAG peptide, which is suitable for immobilization of the proteins to beads containing a cognate affinity partner, here exemplified by the BioM5 anti-FLAG monoclonal antibody.

Table 2. Measured scintillation signals (accumulated under 1 min) from native streptavidin (SA) beads or streptavidin beads coated with biotinylated anti-FLAG BioM5 antibody, respectively, after mixing (and subsequent washing) with transcription/translation mixtures from different samples.

Beads	**Transcription/ translation mix**	**Signal (cpm)**
native SA beads	FLAG-Z_{wt}	4 858
BioM5 anti-FLAG coated	FLAG-Z_{wt}	34 966
native SA beads	FLAG-Z_{IgA}	5 959
BioM5 anti-FLAG coated	FLAG-Z_{IgA}	43 727

Example 5

Cell free transcription/translation of a FLAG-Z_{wt} encoding PCR product, biospecific immobilization of the gene product onto beads and analysis by fluorescence-activated cell sorting (FACS)

Cell free transcription and translation of a PCR product obtained by PCR amplification with primers NOOL-12 and NOOL-13 (Table 1) on a pGEM-SD-K-FLAG-Z_{wt} plasmid template was performed as in example 3, but without the addition of [^{35}S]methionine. The resulting mixture was incubated for 2 hours with 50 mg streptavidin-coated polystyrene beads with a diameter of approximately 0.95 mm) (Bangs Laboratories, Fishers, IN, USA), previously incubated with 5 μl of a 1.89 mg/ml solution of biotinylated anti-FLAG BioM5 monoclonal antibodies. In

the experiment, beads not coated with the biotinyalted BioM5 anti-FLAG antibody were also included, as a control. After thorough washing with TNT buffer (0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.05% Tween 20), rabbit anti-DNP IgG antibodies conjugated to horse-radish peroxidase (HRP) (art. no. P0402, Dako, Denmark) were added to the beads and incubated for 45 min at 25°C, followed by washing with TNT buffer, to detect the translated and biospecifically immobilized FLAG-Z_{wt} fusion protein gene product via the biospecific interaction between the constant parts (Fc) of the rabbit antibodies and the Z domain moiety of the fusion protein. To obtain a signal useful for FACS, the enzymatic activity of the HRP conjugated to the rabbit antibodies was used through the addition of one ml of a signal amplification mixture containing fluorescein tyramide (Anton et al. J. Histochem. Cytochem. 46:771-777, 1998). Between each incubation step the beads were thoroughly washed, centrifuged for 3 min at 2000 x g followed by resuspension in TNT buffer (0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.05% Tween 20) to remove non-specifically bound protein. TNB blocking buffer (0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.5% Blocking reagent from Tyramide Signal Amplification kit, NEN Life Science, Boston, MA, USA) was used during the incubation steps according to the manufacturers instructions.

After an incubation for five minutes at 25°C, and subsequent washing, the beads were resuspended in PBS for FACS analysis. This analysis showed that beads coated with the biotinylated BioM5 anti-FLAG antibody, incubated with the transcription/translation mixture of the FLAG-Z_{wt} encoding PCR product could, subsequently incubated with the rabbit anti-DNP IgG-HRP conjugate and finally subjected to the signal amplification mixture containing fluorescein tyramide displayed significantly higher fluorescence signals in the FACS analysis than

beads treated in the same way, but not containing the BioM5 anti-FLAG antibody (Figure 11).

This shows that fusion proteins, here exemplified by the fusion protein FLAG-Z_{wt}, can be produced from a corresponding PCR product by cell free transcription/ translation containing a functional affinity fusion partner, here exemplified by the FLAG peptide, which is capable of resulting in a biospecific immobilization of the protein to beads containing a cognate affinity partner, here exemplified by the BioM5 anti-FLAG monoclonal antibody, and that such beads can be detected by FACS analysis using a suitable combination of detection reagents, here exemplified by a rabbit anti-DNP IgG-HRP conjugate and a signal amplification mixture containing fluorescein tyramide.

Example 6

Cell free transcription/translation of a bead-immobilized FLAG-Z_{wt} encoding PCR product, biospecific immobilization of the gene product onto beads and analysis by fluorescence-activated cell sorting (FACS)

Biotinylated PCR fragments encoding a FLAG-Z_{wt} fusion protein, obtained after PCR amplification using primers NOOL-12 and NOOL-13 on a plasmid pGEM-SD-K-FLAG-Z_{wt} template were immobilized on streptavidin-coated beads (Bangs Laboratories) at a concentration of approximately 10 ng/mg beads. The beads (50 mg) had previously been incubated with 5 μl of a solution containing 1.89 mg/ml of a biotinylated anti-FLAG peptide antibody (BioM5, Sigma). The beads containing both the biotinylated PCR products and the anti-FLAG peptide antibody were subjected to cell free transcription and translation using 25 ml of an S30 extract (Promega, Madison, WI, USA), supplemented with 200 units of T7 RNA polymerase

(Epicentre, Madison, WI, USA) and 40 units of rRNasin (Promega, Madison, WI, USA). After incubation for one hour at 25°C, followed by repeated washing using TNT buffer (0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.05% Tween 20), rabbit anti-DNP IgG antibodies conjugated to horse-radish peroxidase (HRP) (art. no. P0402, Dako, Denmark) were added to the beads and incubated overnight at 4°C (end-over-end mixing), followed by washing with TNT, to detect the translated and biospecifically immobilized FLAG-Z_{wt} fusion protein gene product via the biospecific interaction between the constant parts (Fc) of the rabbit antibodies and the Z domain moiety of the fusion protein (Nilsson et al., Protein engineering, 1: 107-113, 1987).

To obtain a signal useful for FACS, the enzymatic activity of the HRP conjugated to the rabbit antibodies was used through the addition of one ml of a signal amplification mixture containing fluorescein tyramide (Anton et al. J. Histochem. Cytochem. 46:771-777, 1998). Between each incubation step the beads were thoroughly washed, centrifuged for 3 min at 2000 x g followed by resuspension in TNT buffer (0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.05% Tween 20) to remove non-specifically bound protein. TNB blocking buffer (0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.5% Blocking reagent from Tyramide Signal Amplification kit, NEN Life Science, USA) was used during the incubation steps according to the manufacturers instructions. As a negative control, streptavidin coated beads, containing immobilized BioM5 anti-FLAG antibodies, and a PCR products obtained from PCR amplification using primers NOOL-12 and NOOL-13 on a plasmid pGEM-SD-K-FLAG-Z_{IgA} template were included in the experiment.

The results from the FACS analysis shows that the beads containing the immobilized biotinylated PCR fragments

encoding a FLAG-Z_{wt} fusion protein, obtained after PCR amplification using primers NOOL-12 and NOO1-13 on a plasmid pGEM-SD-K-FLAG-Z_{wt} template display a significantly higher fluorescence intensity than the control beads containing immobilized PCR products encoding a fusion protein not recognized by the reagent rabbit-HRP conjugate used for detection (Figure 12). This shows that fusion proteins, here exemplified by the fusion protein FLAG-Z_{wt}, can be produced from a corresponding, bead-immobilized, PCR product by cell free transcription/translation, containing a functional affinity fusion partner, here exemplified by the FLAG peptide, which is capable of resulting in a biospecific immobilization of the protein to beads containing a cognate affinity partner, here exemplified by the BioM5 anti-FLAG monoclonal antibody, and that such beads can be detected by FACS analysis using a suitable combination of detection reagents, here exemplified by a rabbit anti-DNP IgG-HRP conjugate and a signal amplification mixture containing fluorescein tyramide.

Example 7

Fluorescence-activated cell sorting (FACS)-based enrichment of beads containing immobilized PCR products encoding a desired gene product

Biotinylated PCR fragments encoding FLAG-Z_{wt} and FLAG-Z_{IgA} fusion proteins, respectively, obtained after PCR amplification using primers NOOL-12 and NOOL-13 on plasmids pGEM-SD-K-FLAG-Z_{wt} and pGEM-SD-K-FLAG-Z_{IgA} templates, respectively were separately immobilized on streptavidin-coated beads (Bangs Laboratories) to a level of approximately 10 ng/mg beads. The beads (50 mg) had previously been incubated with 5 μl of a solution containing 1.89 mg/ml of a biotinylated anti-FLAG peptide antibody (BioM5, Sigma, Saint Louis, Mo, USA). Beads from the two pools were subsequently

mixed at a ratio of 1:1 (equal amounts of beads of both sorts) and subjected to cell free transcription and translation using 25 ml of an S30 extract (Promega, Madison, WI, USA), supplemented with 200 units of T7 RNA polymerase (Epicentre) and 40 units of rRNasin (Promega). After incubation for one hour at 25°C, followed by repeated washing using TNT buffer (0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.05% Tween 20), rabbit anti-DNP IgG antibodies conjugated to horse-radish peroxidase (HRP) (art. no. P0402, Dako, Denmark) were added to the beads and incubated for overnight at 4°C, followed by washing with TNT, to detect the translated and biospecifically immobilized FLAG-Z_{wt} fusion protein gene product via the biospecific interaction between the constant parts (Fc) of the rabbit antibodies and the Z domain moiety of the fusion protein. To obtain a signal useful for FACS, the enzymatic activity of the HRP conjugated to the rabbit antibodies was used through the addition of one ml of a signal amplification mixture containing fluorescein tyramide (Anton et al. J. Histochem. Cytochem. 46:771-777, 1998). Between each incubation step the beads were thoroughly washed, centrifuged for 3 min at 2000 x g followed by resuspension in TNT buffer (0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.05% Tween 20) to remove non-specifically bound protein. TNB blocking buffer (0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.5% Blocking reagent from Tyramide Signal Amplification kit, NEN Life Science, Boston, MA, USA) was used during the incubation steps according to the manufacturers instructions. Using FACS, a bead pool originally obtained by the mixing at the 1:1 bead ratio was subsequently subjected to enrichment experiment based on fluorescence intensity. In this procedure the settings in the FACS instrument were adjusted for preparative isolation of single beads (singlets) having a relative fluorescence intensity above 50. With this setting, the mixture was subjected to sorting and tubes

with approximately 4500 sorted beads were collected.

To analyze if beads carrying the PCR products encoding the FLAG-Z_{wt} fusion protein, which should be specifically labeled by the labeling procedure involving the rabbit IgG-HRP conjugate, were enriched relative to beads carrying the PCR products and FLAG-Z_{IgA} fusion proteins not being recognized by the rabbit IgG-HRP conjugate, the difference in DNA sequence between the two PCR products was employed.

The FLAG-Z_{wt} fusion protein-encoding PCR products contain a recognition sequence for the enzyme Mlu I, not present in the PCR products encoding the FLAG-ZIgA fusion protein. This allowed for a discrimination between the two PCR products through an analysis of the susceptibility for Mlu I digestion (Figure 13A). Samples of beads from before and after sorting were therefore subjected to PCR amplification using primers NOOL-12 and NOOL-13, which anneals at sites in the immobilized PCR products flanking the regions which differs between the two PCR product species, and therefore could be use for the simultaneous amplification of both PCR product species. Subsequent incubation of the resulting new PCR products with the restriction enzyme Mlu I could therefore be used to investigate the relative ratios between the two species in samples from before and after sorting, by analysis of DNA fragment sizes and band intensities after agarose gel electrophoresis followed by ethidium bromide staining.

A PCR amplification of the nucleic acids present on approximately 10000 beads from the 1:1 mixture (sample from before sorting) followed by a digestion with Mlu I and analysis by gel electrophoresis shows, as expected, upon a mixture of Mlu I-susceptible and Mlu I-resistent

PCR products (Figure 13B, lane 6).

When approximately 400 beads collected during the FACS enrichment was subjected to the same analysis, the intensity ratio between the upper band (443 bp, uncleaved) and lower double band (two cleavage products, 239/204 bp, unresolved) had shifted towards the smaller (lower) bands (Figure 13B, lane 8). Using a Gel Doc 2000 gel scanning instrument and Quantity One vers. 4.1 software (Biorad, Hercules, CA, USA), this shift in relative intensities were recorded resulting in the overlay plot shown in figure 14. From this analysis it can be clearly seen that a shift of the relative intensity towards the lower molecular weight cleavage products had occured. This shows that beads containing Mlu I-susceptible PCR product encoding the FLAG-Z_{wt} fusion protein, had been enriched during the experiment, relative to beads containing the Mlu I-resistent FLAG-Z_{IgA} fusion protein encoding PCR product.

Taken together, this example shows that fusion proteins, here exemplified by the fusion protein FLAG-Z_{wt}, can be produced from a corresponding, bead-immobilized, PCR product by cell free transcription/translation, containing a functional affinity fusion partner, here exemplified by the FLAG peptide, which is capable of resulting in a biospecific immobilization of the protein to beads containing a cognate affinity partner, here exemplified by the BioM5 anti-FLAG monoclonal antibody, and that such beads can be enriched when mixed and co-processed with irrelevant beads, containing PCR products encoding a different gene product, by FACS-based enrichment using a suitable combination of detection reagents, here exemplified by a rabbit anti-DNP IgG-HRP conjugate and a signal amplification mixture containing fluorescein tyramide.

Claims

1. A method for the selection of one or more desired polypeptides comprising:

(a) cell free expression of nucleic acid molecules immobilized on a solid support system to produce polypeptides, the solid support carrying means for biospecific interaction with at least the desired polypeptide or a molecule attached thereto;

(b) separation of the solid support carrying both the desired polypeptide and the nucleic acid encoding it; and optionally

(c) recovery of the said nucleic acid and/or said desired polypeptide.

2. A method as claimed in claim 1 wherein the expressed polypeptides are fusion proteins.

3. A method as claimed in claim 2 wherein each fusion protein comprise a variable portion and a common portion.

4. A method as claimed in claim 3 wherein the common portion comprises an affinity fusion partner whose cognate binding partner is immobilised on the solid support.

5. A method as claimed in claim 3 wherein the common portion comprises a reporter protein moiety.

6. A method as claimed in any one of claims 3 to 5 wherein the variable portion is a member of a polypeptide library.

7. A method as claimed in any one of the preceding claims wherein steps (a) and (b) are performed iteratively for more than one cycle.

8. A method as claimed in claim 7 wherein steps (a) and (b) are performed between 2 and 20 times.

9. A method as claimed in any one of the preceding claims wherein the solid support system is particulate.

10. A method as claimed in claim 9 wherein immobilised on each solid support particle is a nucleic acid molecule and said means for biospecific interaction with at least the desired polypeptide or a molecule attached thereto.

11. A method as claimed in any preceding claim wherein the immobilised means for biospecific interaction is a target molecule for the desired polypeptide.

12. A method as claimed in any one of claims 1 to 10 wherein the immobilised means for biospecific interaction is a cognate binding partner for an affinity binding partner which forms a fusion protein with the desired polypeptide.

13. A nucleic acid molecule or polypeptide when selected according to the method of any preceding claim.

14. A molecular library comprising a solid support system having immobilised thereon a plurality of nucleic acid molecules and associated with each of said nucleic acid molecules and also immobilised on said support system means for biospecific interaction with the expression product of one or more of said nucleic acid molecules.

15. A library as claimed in claim 14 wherein the solid support system is particulate.

16. A library as claimed in claim 15 wherein

immobilised on each solid support particle is a nucleic acid molecule and means for biospecific interaction with the expression product of one or more of said nucleic acid molecules.

17. A library as claimed in claim 16 wherein the immobilised means for biospecific interaction is a target molecule for the expression product of one or more of said nucleic acid molecules.

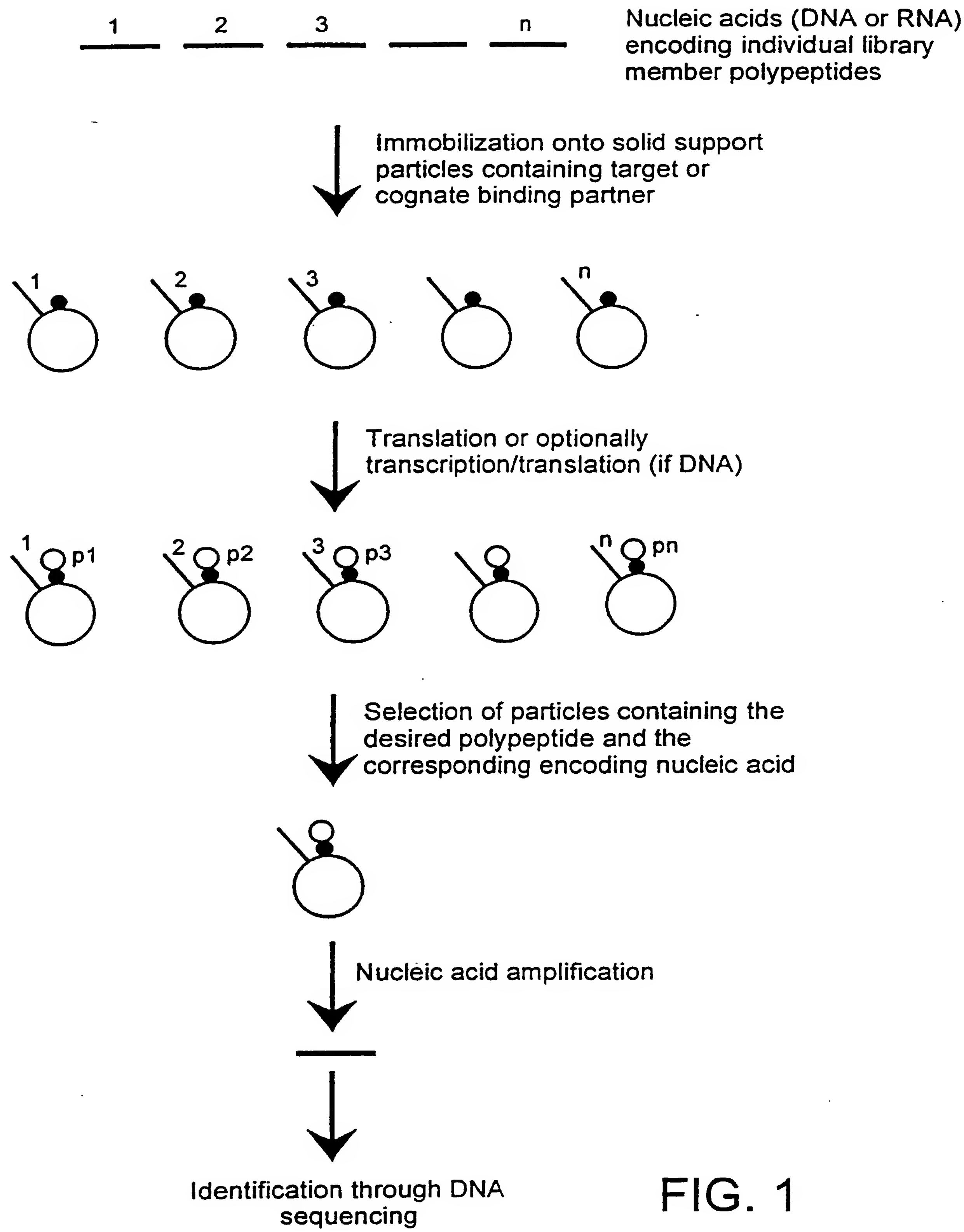

1
2
3
n
Nucleic acids (DNA or RNA) encoding individual library member polypeptides
Immobilization onto solid support particles containing target or cognate binding partner
1
2
3
n
Translation or optionally transcription/translation (if DNA)
1
p1
2
p2
3
p3
n
pn
Selection of particles containing the desired polypeptide and the corresponding encoding nucleic acid
Nucleic acid amplification
Identification through DNA sequencing

FIG. 1

Immobilized DNA/Target added

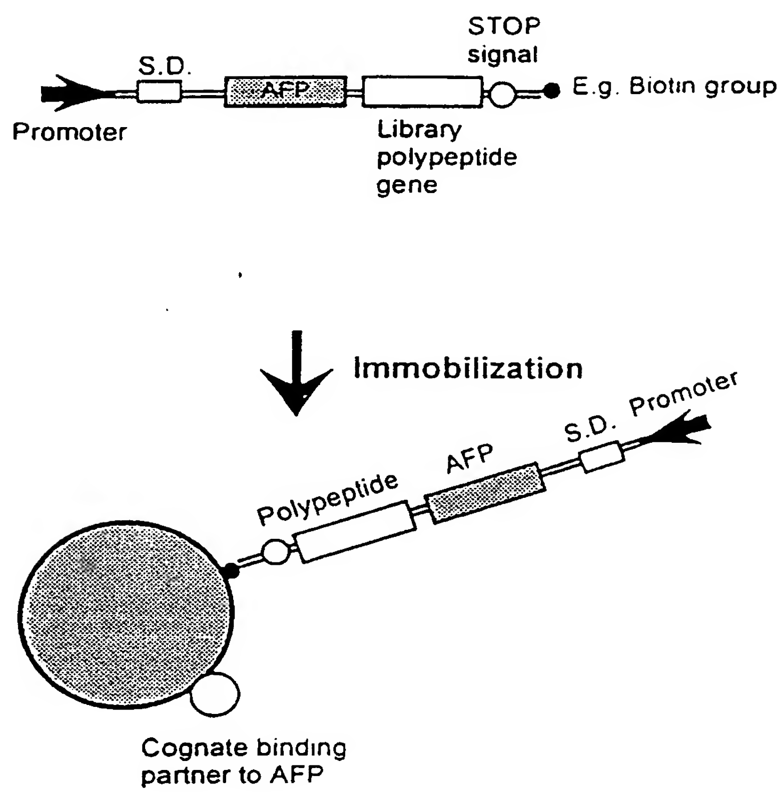


FIG. 2

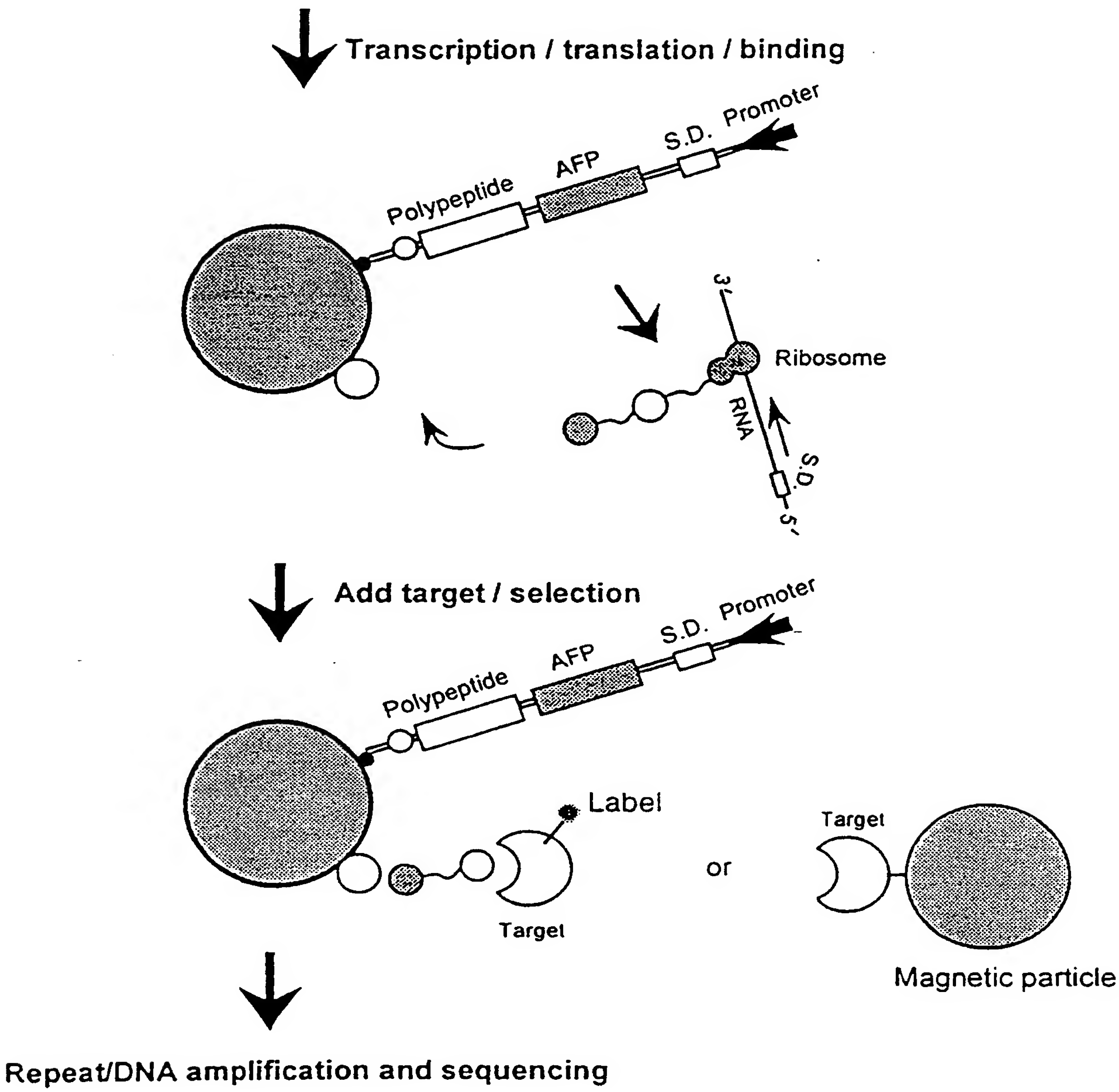

Transcription / translation / binding
Polypeptide
AFP
S.D.
Promoter
3'
Ribosome
RNA
S.D.
5'
Add target / selection
Polypeptide
AFP
S.D.
Promoter
Label
Target
or
Target
Magnetic particle
Repeat/DNA amplification and sequencing

FIG. 2CONT'D

Immobilized RNA/Target added

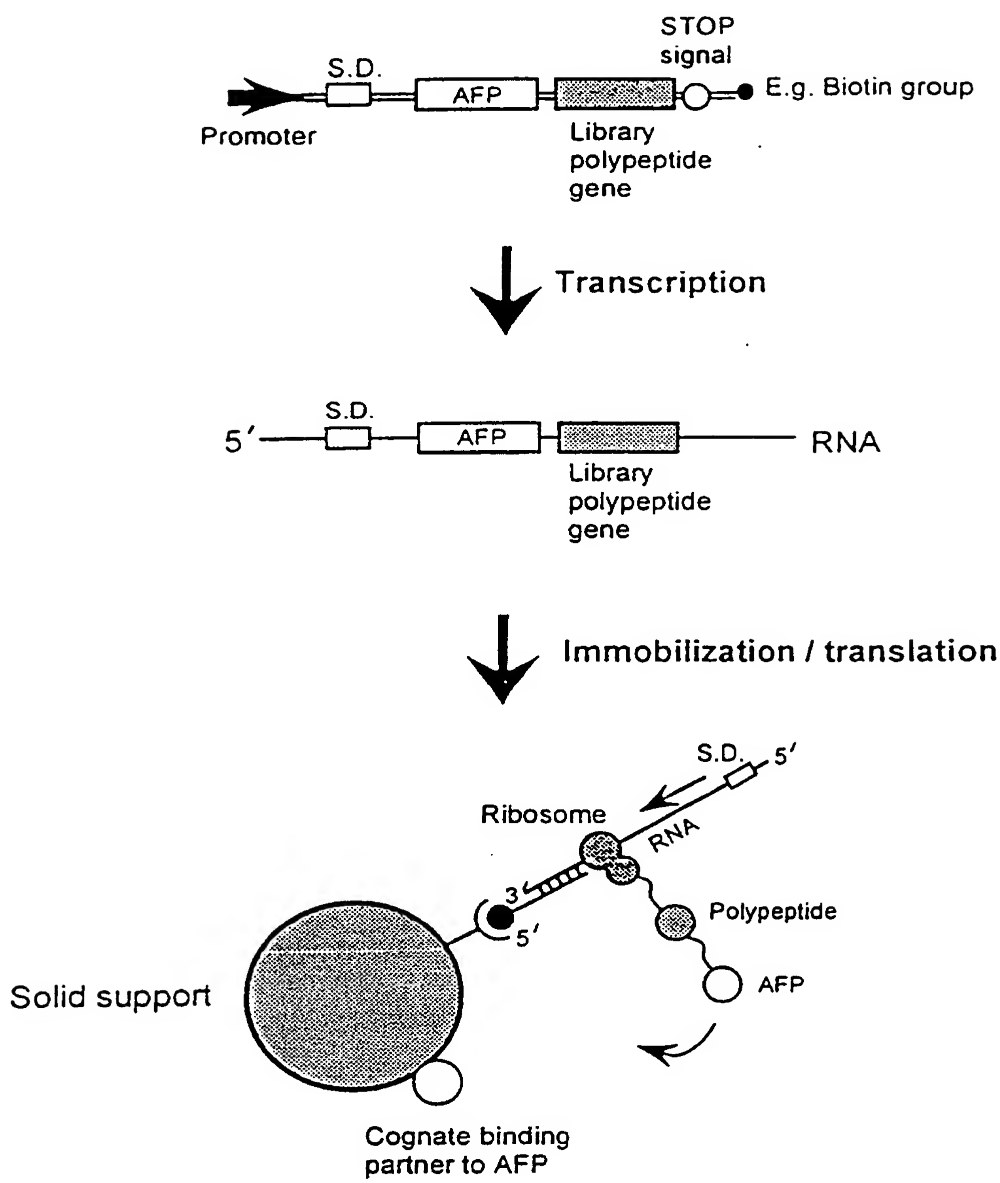
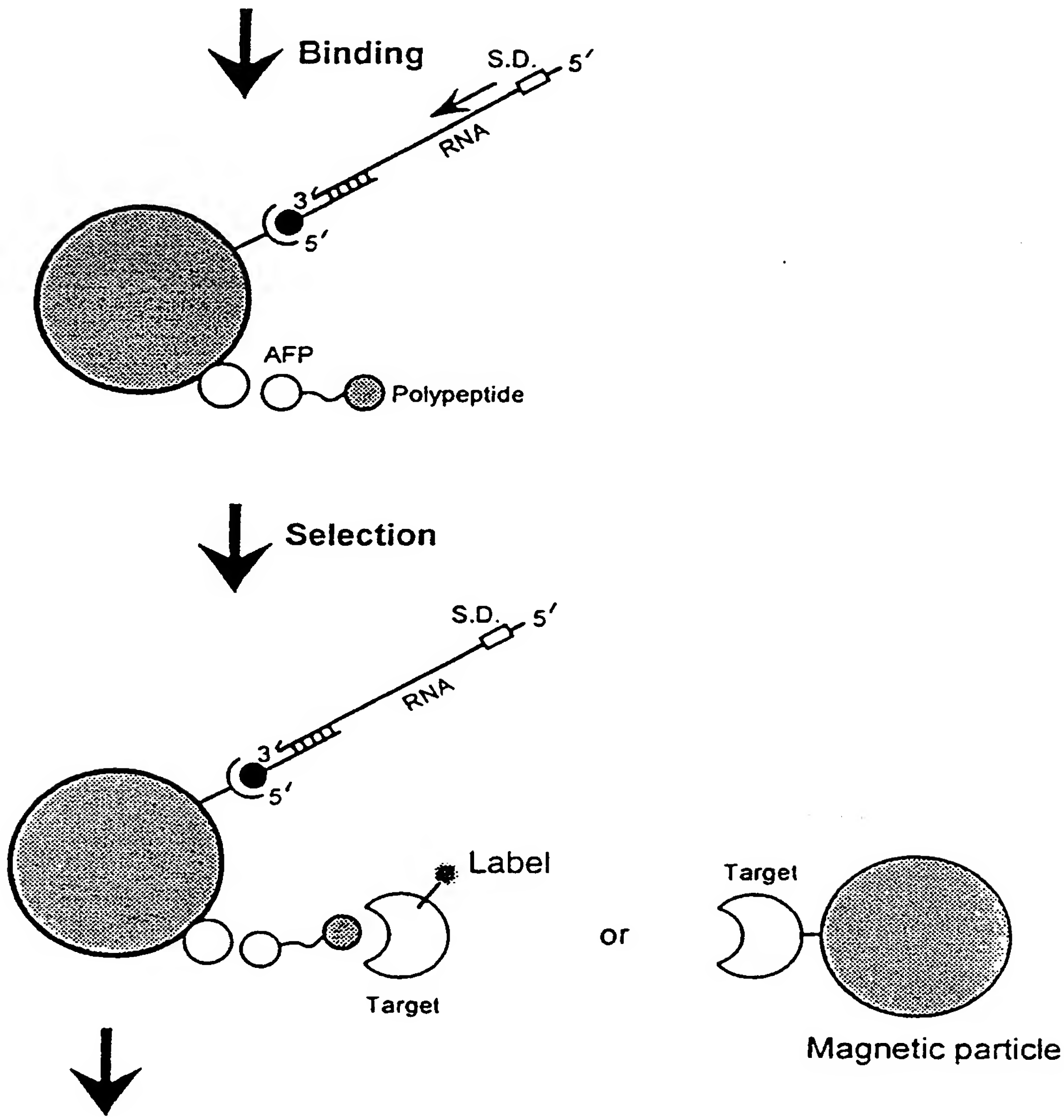


FIG. 3



Reverse transcribe RNA / PCR amplification and sequencing

FIG. 3CONT'D

Immobilized DNA/Target immobilized

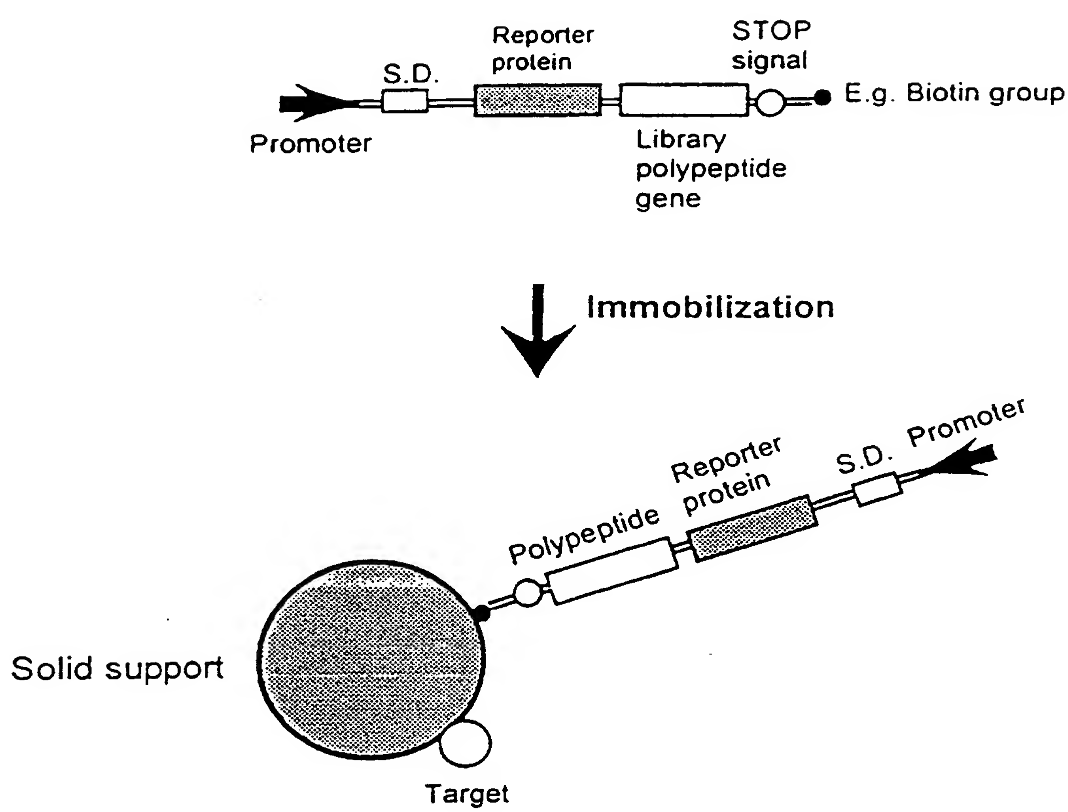


FIG. 4

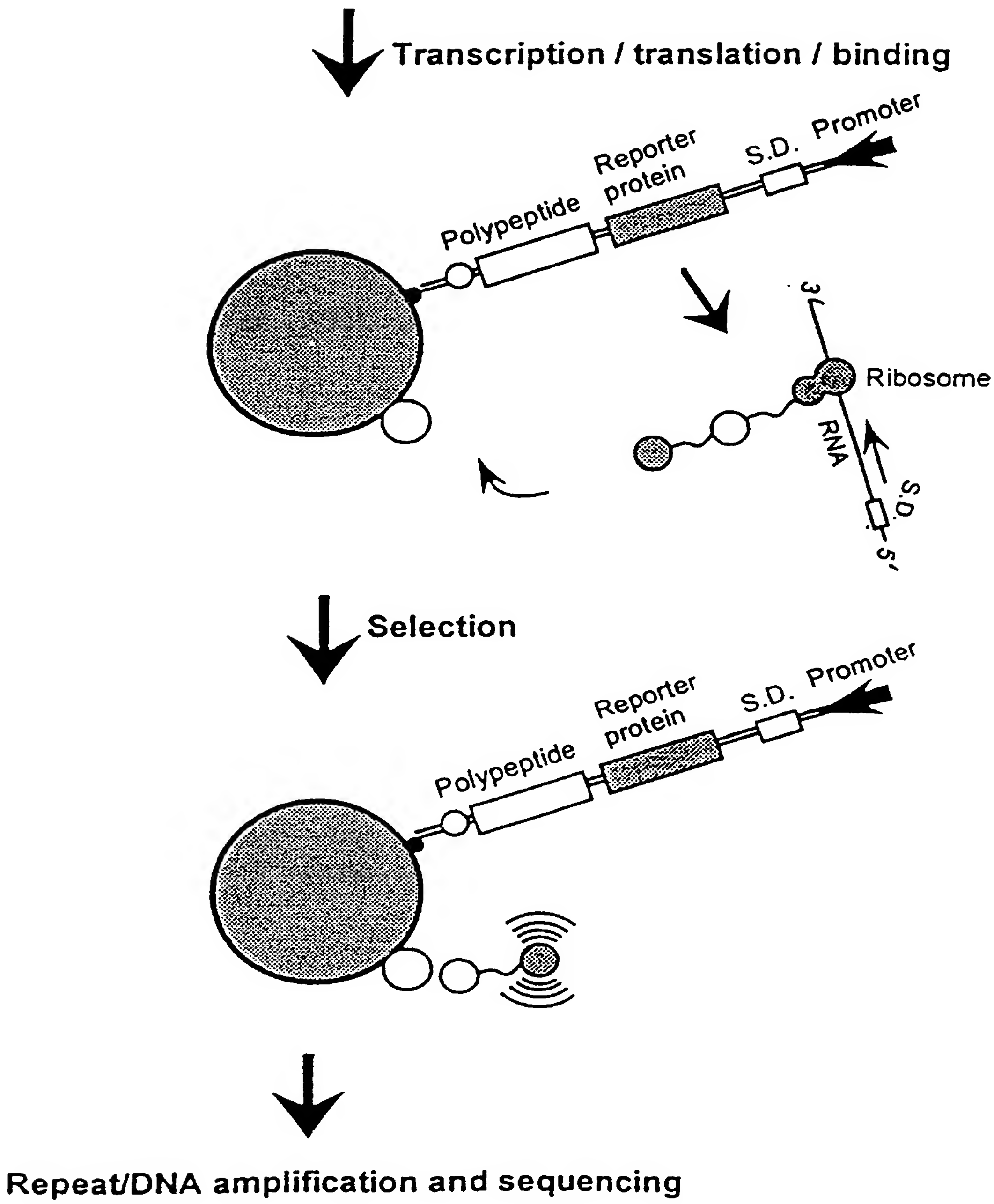

Transcription / translation / binding
Polypeptide
Reporter protein
S.D.
Promoter
3'
Ribosome
RNA
S.D.
5'
Selection
Polypeptide
Reporter protein
S.D.
Promoter
Repeat/DNA amplification and sequencing

FIG. 4CONT'D

Immobilized RNA/Target immobilized

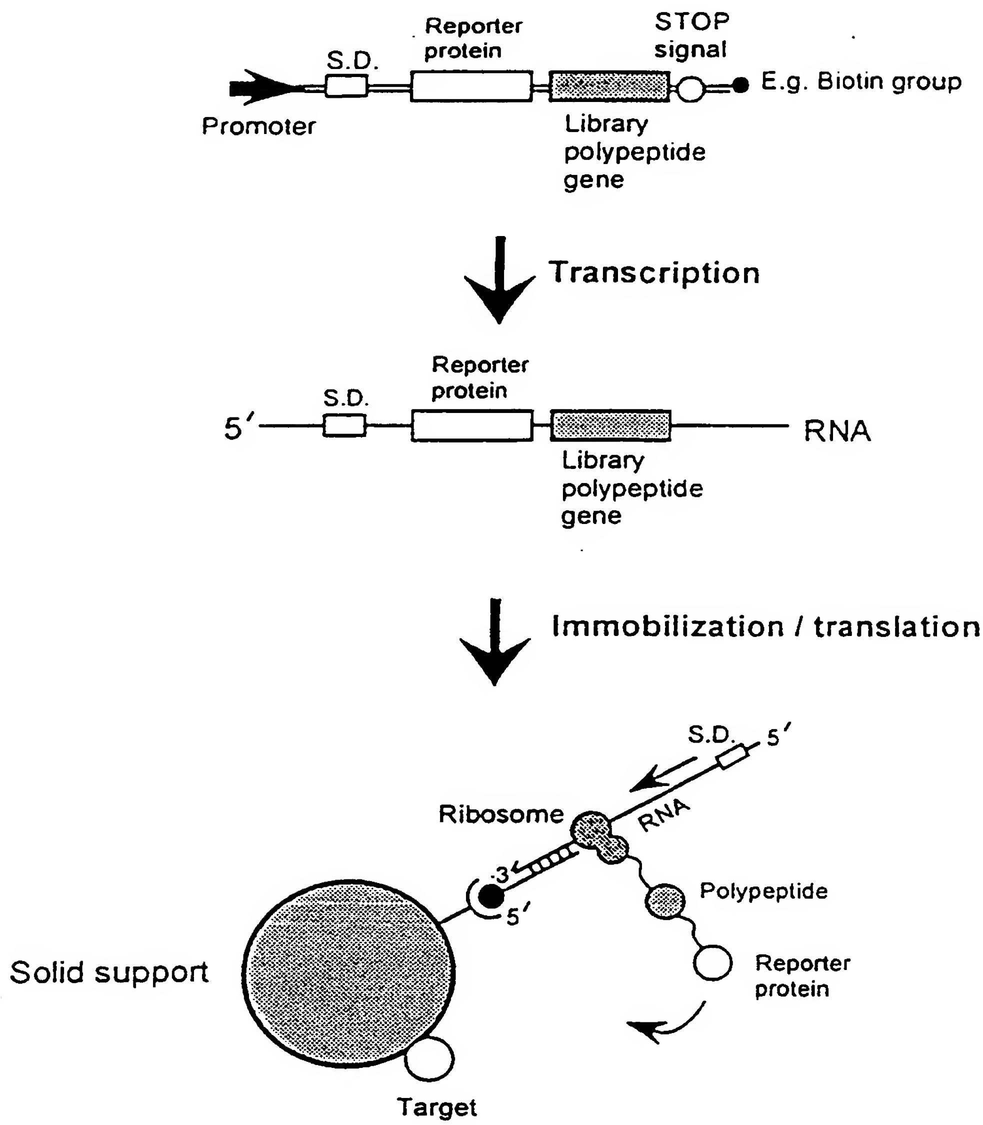
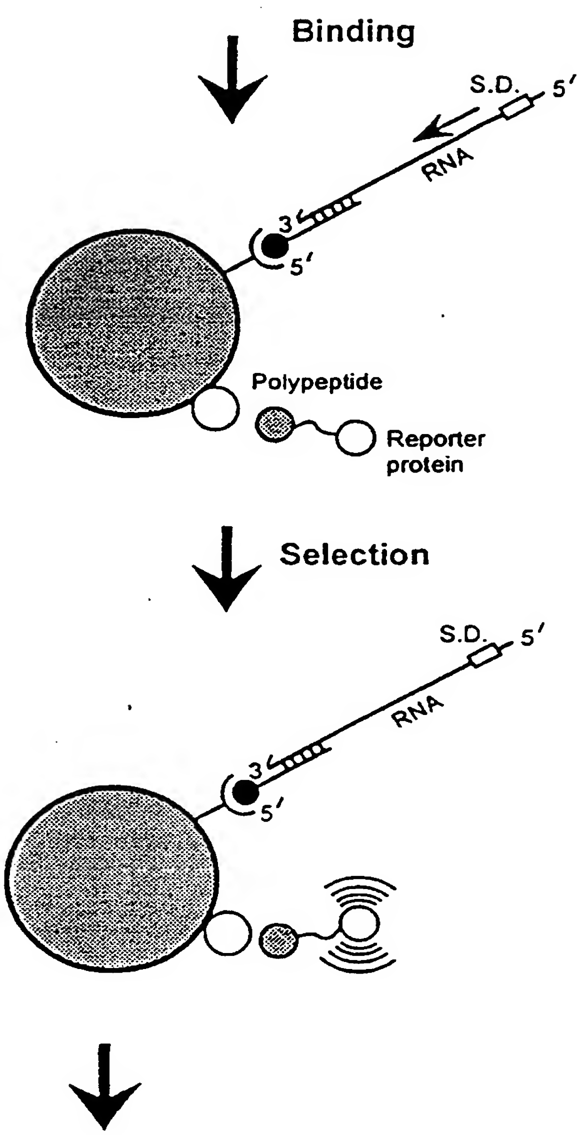


FIG. 5



Reverse transcribe RNA / PCR amplification and sequencing

FIG. 5CONT'D

Example 1.

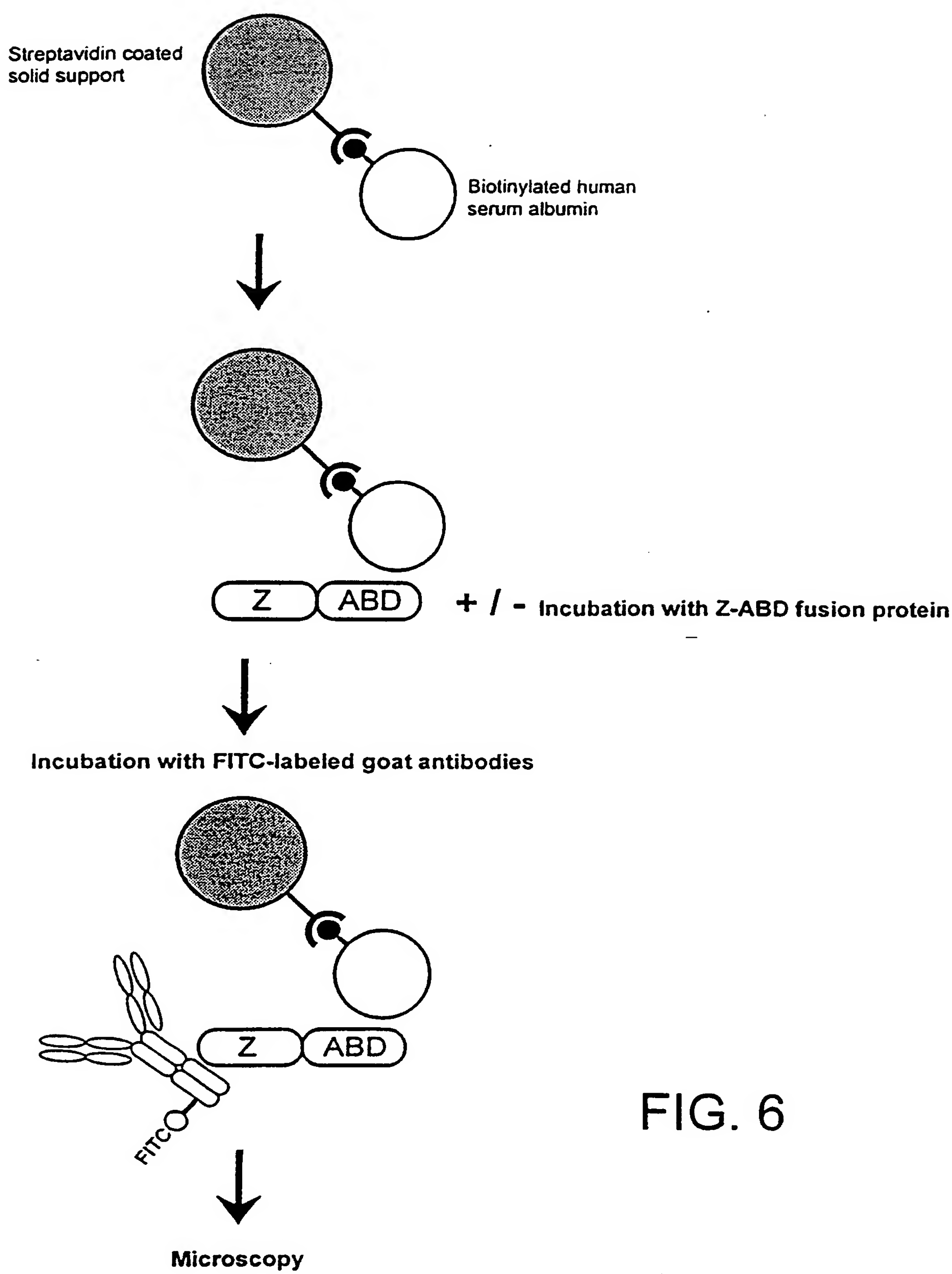


FIG. 6

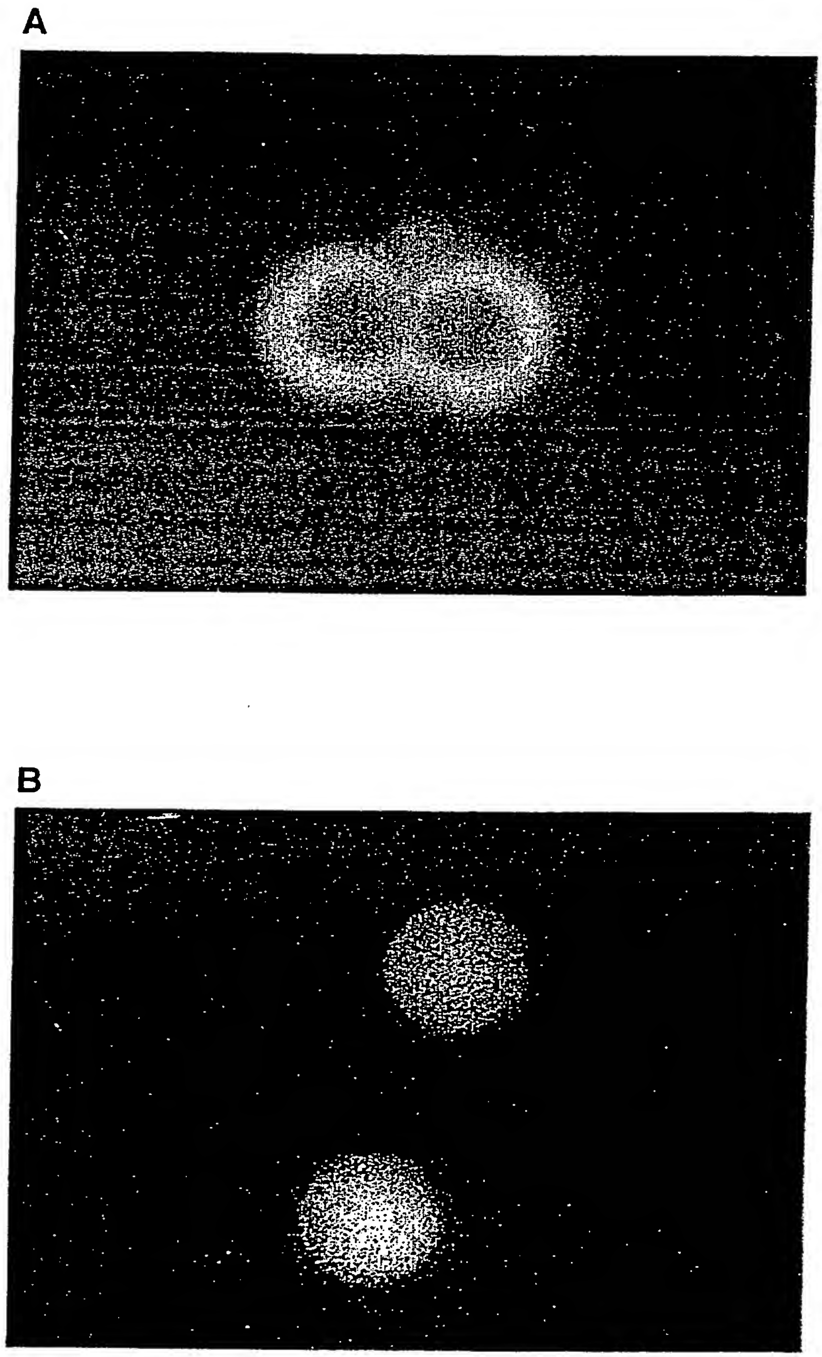


FIG. 7

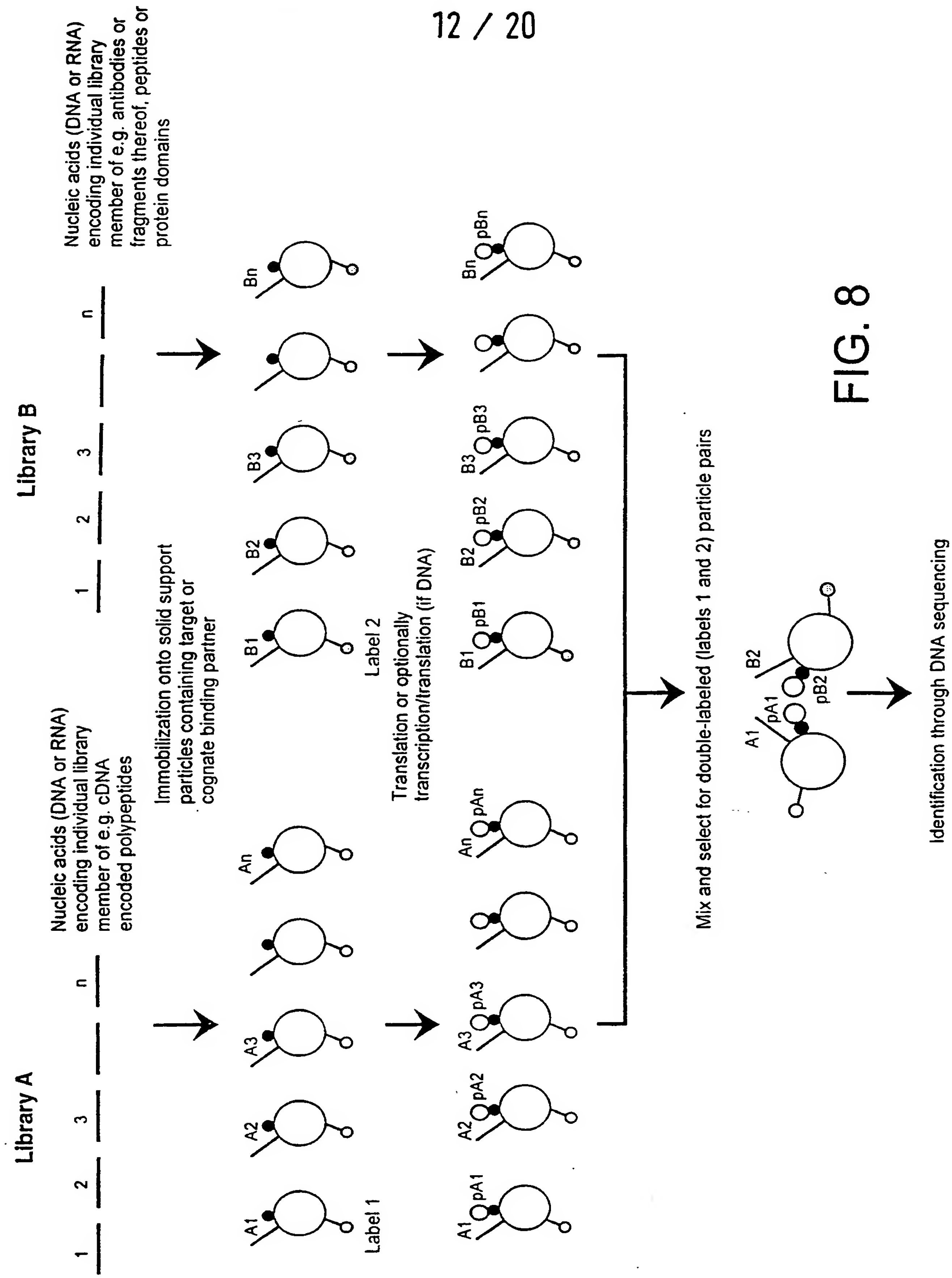
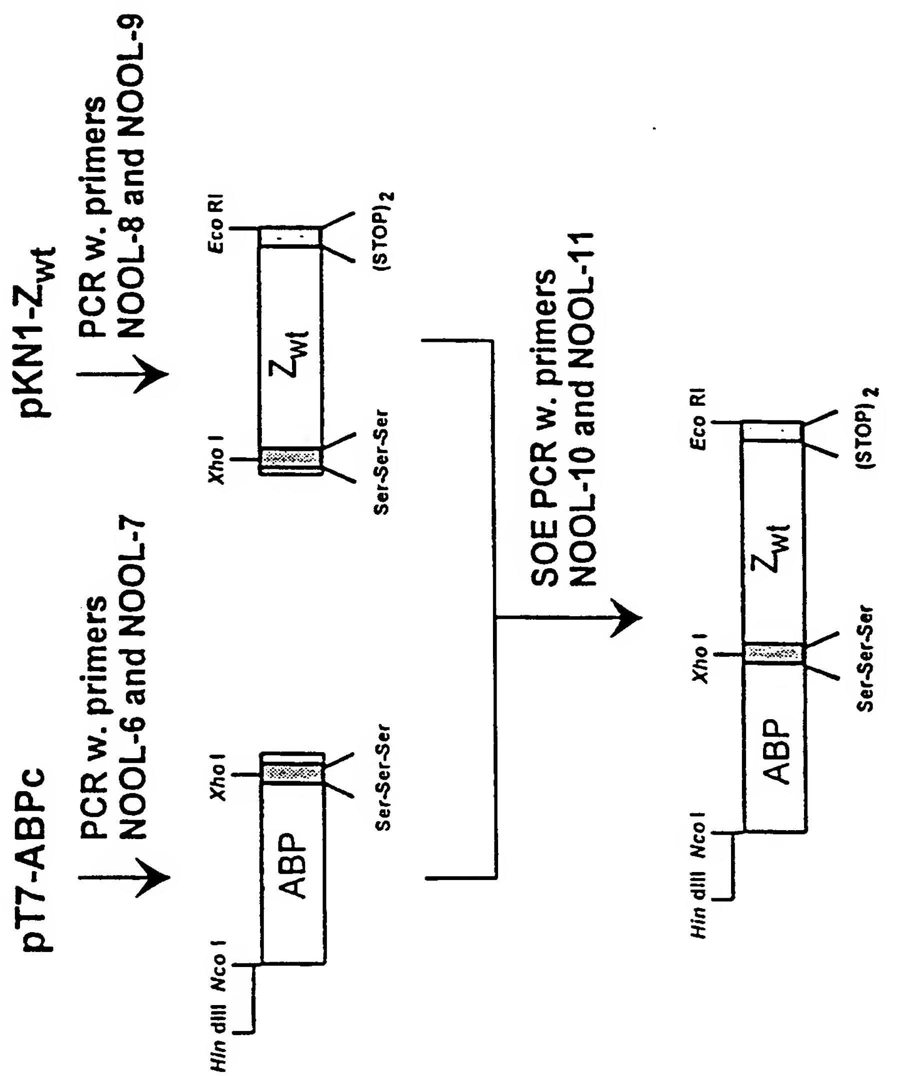

Library A
1
2
3
n
Nucleic acids (DNA or RNA) encoding individual library member of e.g. cDNA encoded polypeptides
Library B
1
2
3
n
Nucleic acids (DNA or RNA) encoding individual library member of e.g. antibodies or fragments thereof, peptides or protein domains
Immobilization onto solid support particles containing target or cognate binding partner
A1
A2
A3
An
B1
B2
B3
Bn
Label 1
Label 2
Translation or optionally transcription/translation (if DNA)
A1 pA1
A2 pA2
A3 pA3
An pAn
B1 pB1
B2 pB2
B3 pB3
Bn pBn
Mix and select for double-labeled (labels 1 and 2) particle pairs
A1
pA1
pB2
B2
Identification through DNA sequencing

FIG. 8

FIG. 9


pT7-ABPc
PCR w. primers NOOL-6 and NOOL-7
Hin dIII
Nco I
ABP
Xho I
Ser-Ser-Ser
pKN1-Z_{wt}
PCR w. primers NOOL-8 and NOOL-9
Xho I
Z_{wt}
Eco RI
Ser-Ser-Ser
$(STOP)_2$
SOE PCR w. primers NOOL-10 and NOOL-11
Hin dIII
Nco I
ABP
Xho I
Z_{wt}
Eco RI
Ser-Ser-Ser
$(STOP)_2$

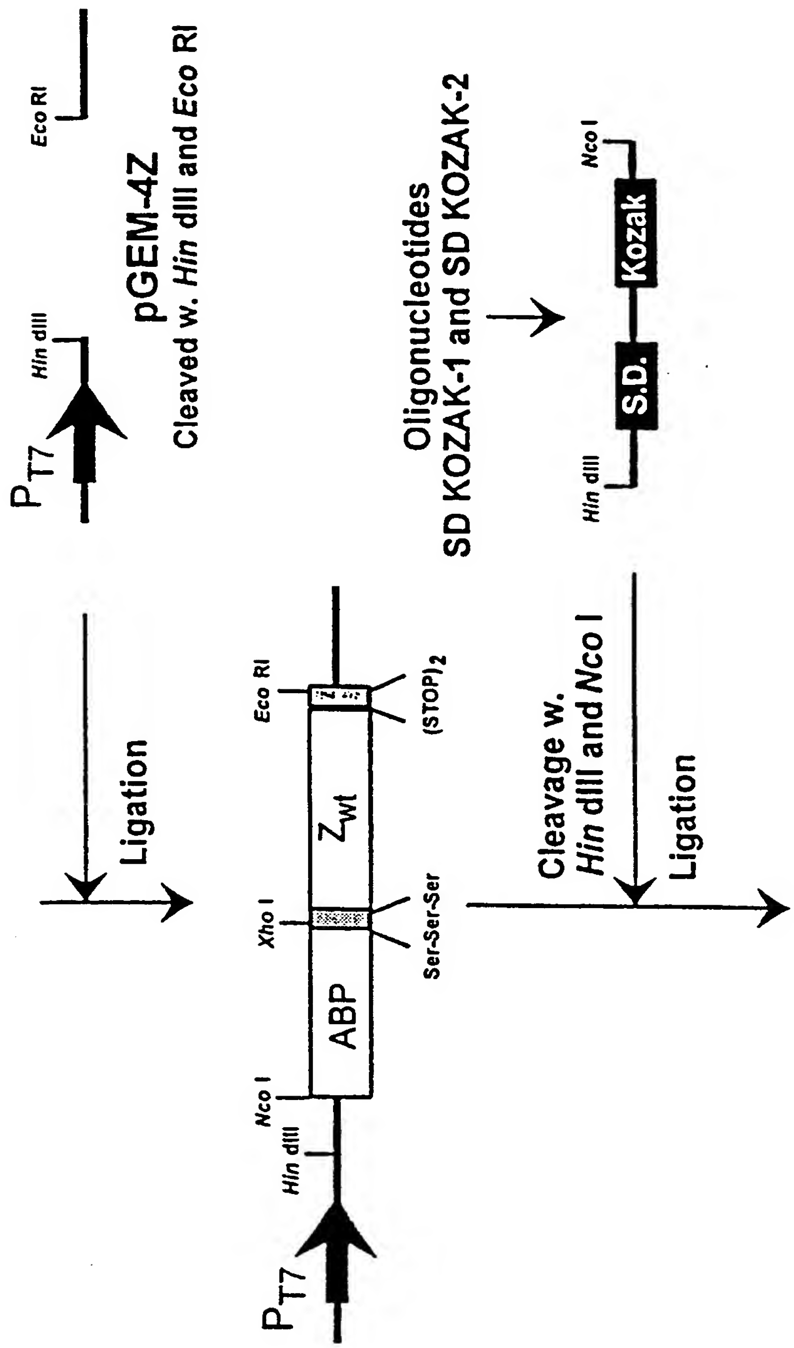

P_T7
Hin dIII
Eco RI
pGEM-4Z
Cleaved w. Hin dIII and Eco RI
Oligonucleotides
SD KOZAK-1 and SD KOZAK-2
Hin dIII
S.D.
Kozak
Nco I
Ligation
P_T7
Hin dIII
Nco I
ABP
Xho I
Ser-Ser-Ser
Z_wt
Eco RI
(STOP)_2
Cleavage w.
Hin dIII and Nco I
Ligation

FIG. 9CONT'D

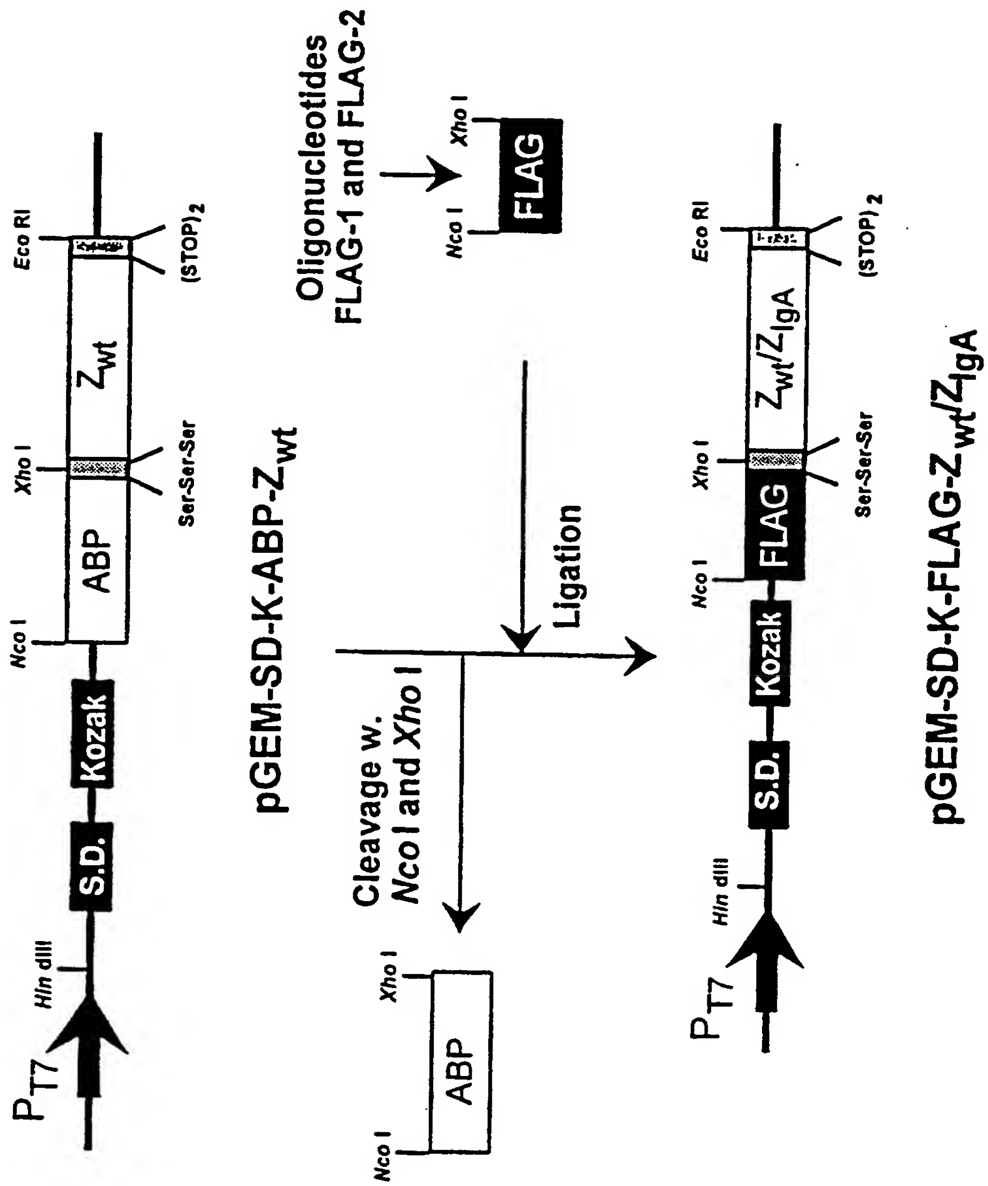

P_{T7}
Hin dIII
S.D.
Kozak
Nco I
ABP
Xho I
Ser-Ser-Ser
Z_{wt}
Eco RI
(STOP)$_2$
pGEM-SD-K-ABP-Z_{wt}
Cleavage w.
Nco I and *Xho* I
Nco I
ABP
Xho I
Oligonucleotides
FLAG-1 and FLAG-2
Nco I
FLAG
Xho I
Ligation
P_{T7}
Hin dIII
S.D.
Kozak
Nco I
FLAG
Xho I
Ser-Ser-Ser
Z_{wt}/Z_{IgA}
Eco RI
(STOP)$_2$
pGEM-SD-K-FLAG-Z_{wt}/Z_{IgA}

FIG. 9CONT'D

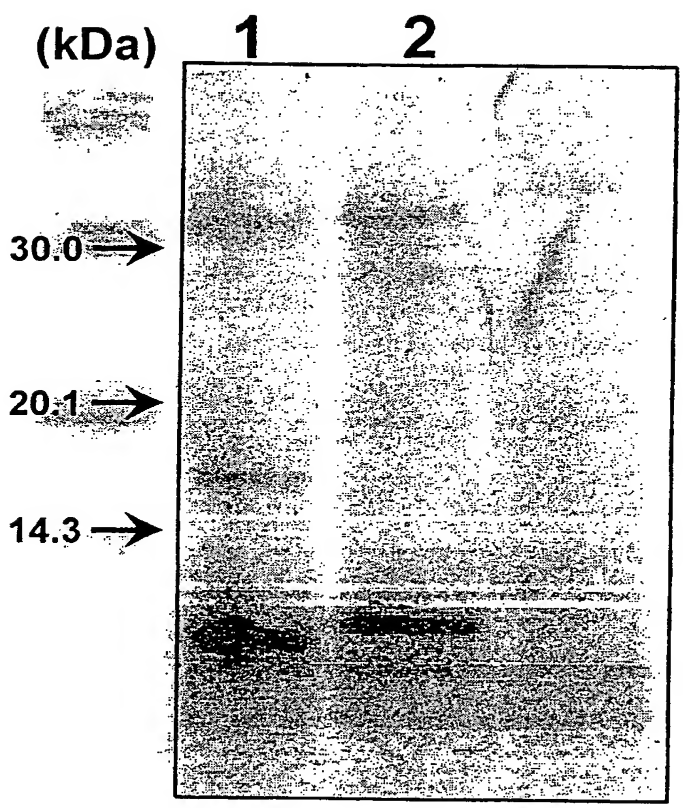


FIG. 10

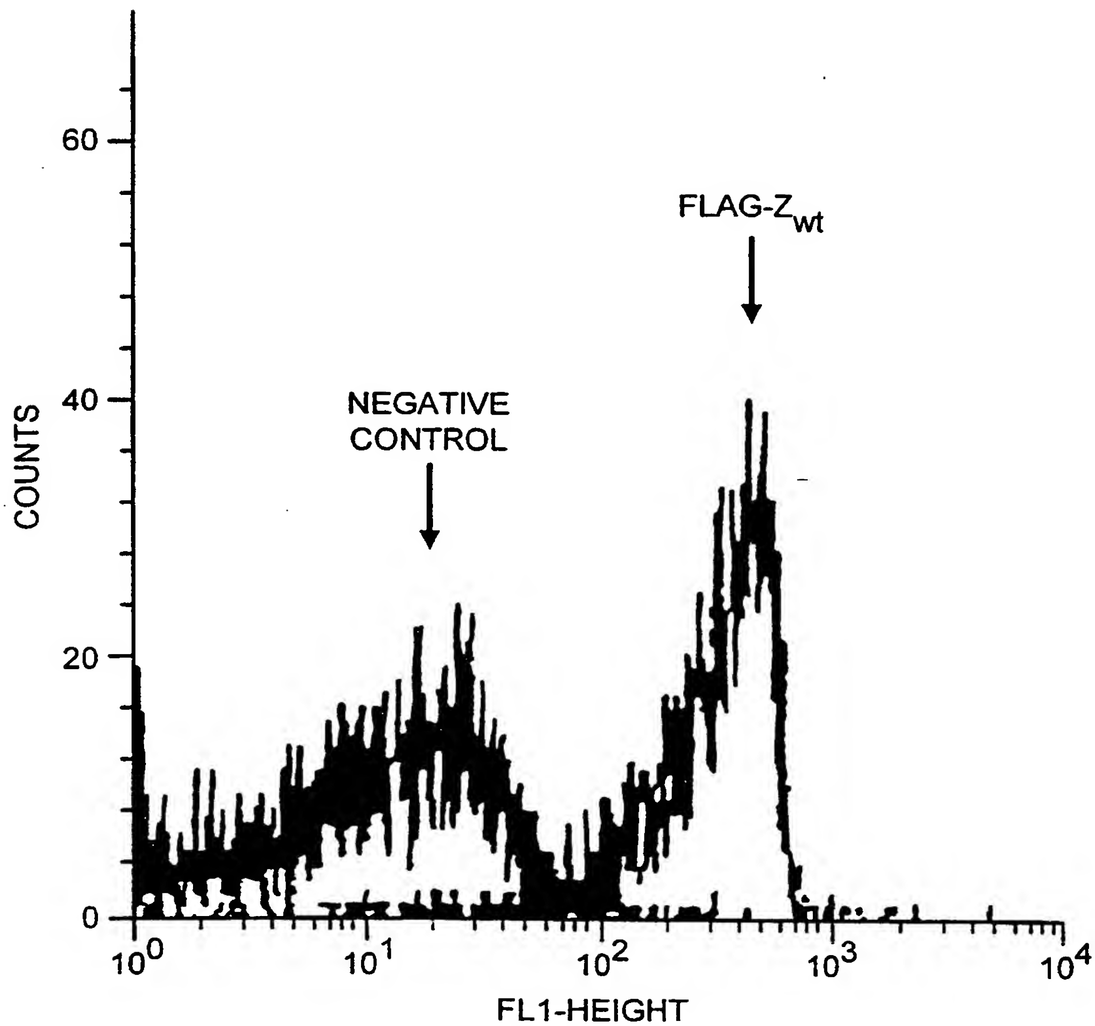

60
40
20
0
COUNTS
FLAG-Z_{wt}
NEGATIVE CONTROL
10^0
10^1
10^2
10^3
10^4
FL1-HEIGHT

FIG. 11

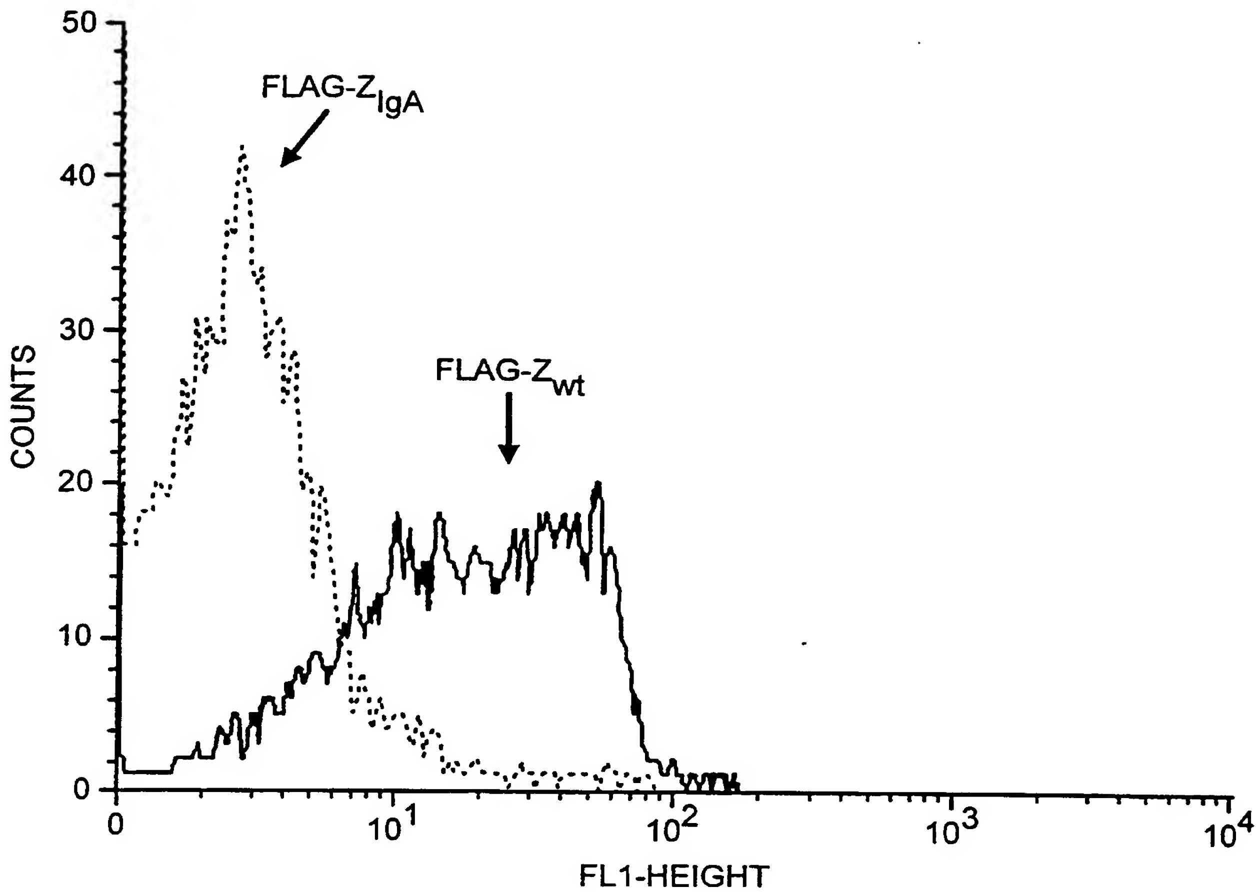

50
40
30
20
10
0
COUNTS
FLAG-Z_{IgA}
FLAG-Z_{wt}
0
10^1
10^2
10^3
10^4
FL1-HEIGHT

FIG. 12

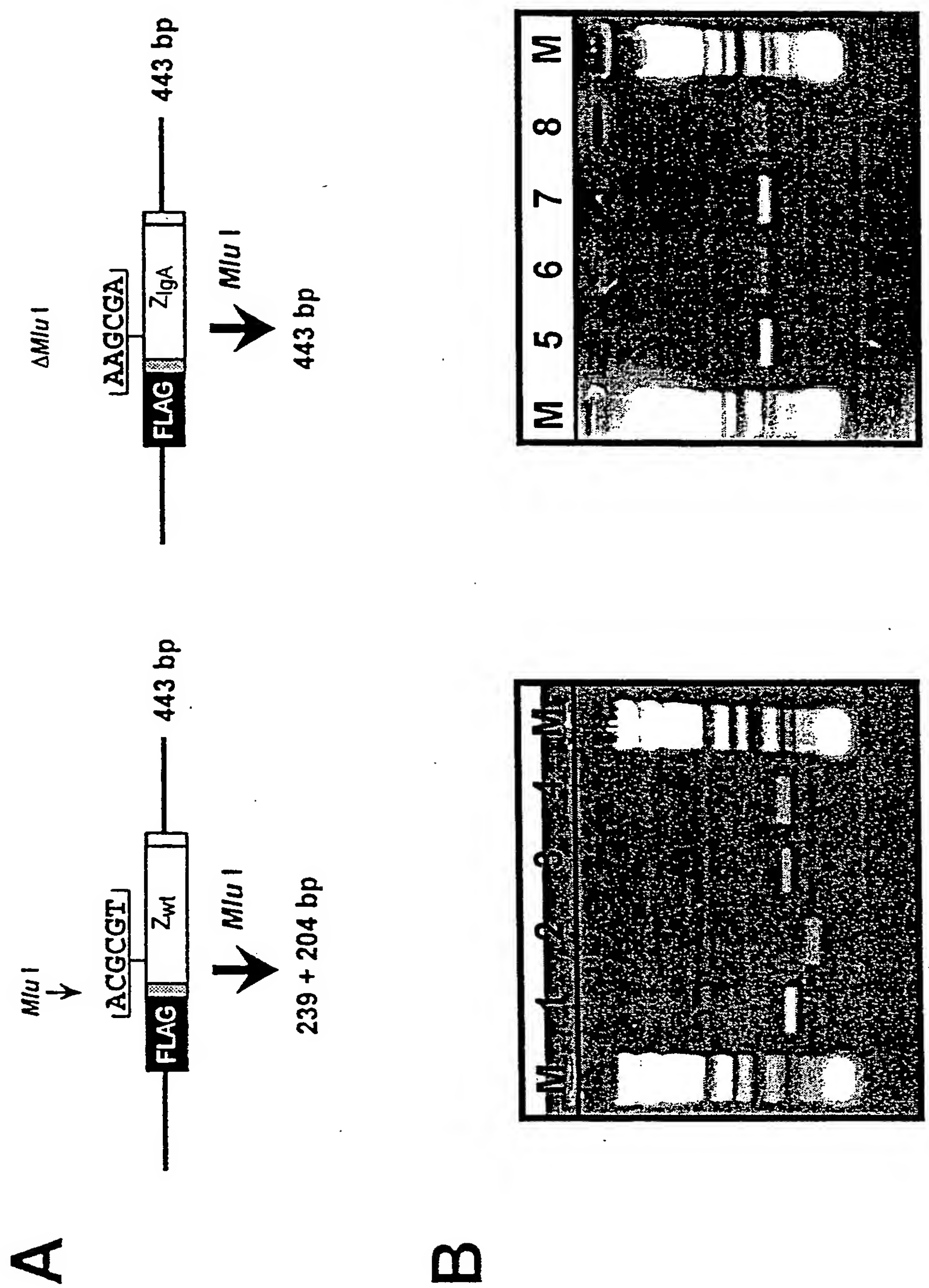

A
Mlu I
ACGCGT
FLAG
Zwt
443 bp
Mlu I
239 + 204 bp
ΔMlu I
AAGCGA
FLAG
ZIgA
443 bp
Mlu I
443 bp
B
M 1 2 3 4 M
M 5 6 7 8 M

FIG. 13

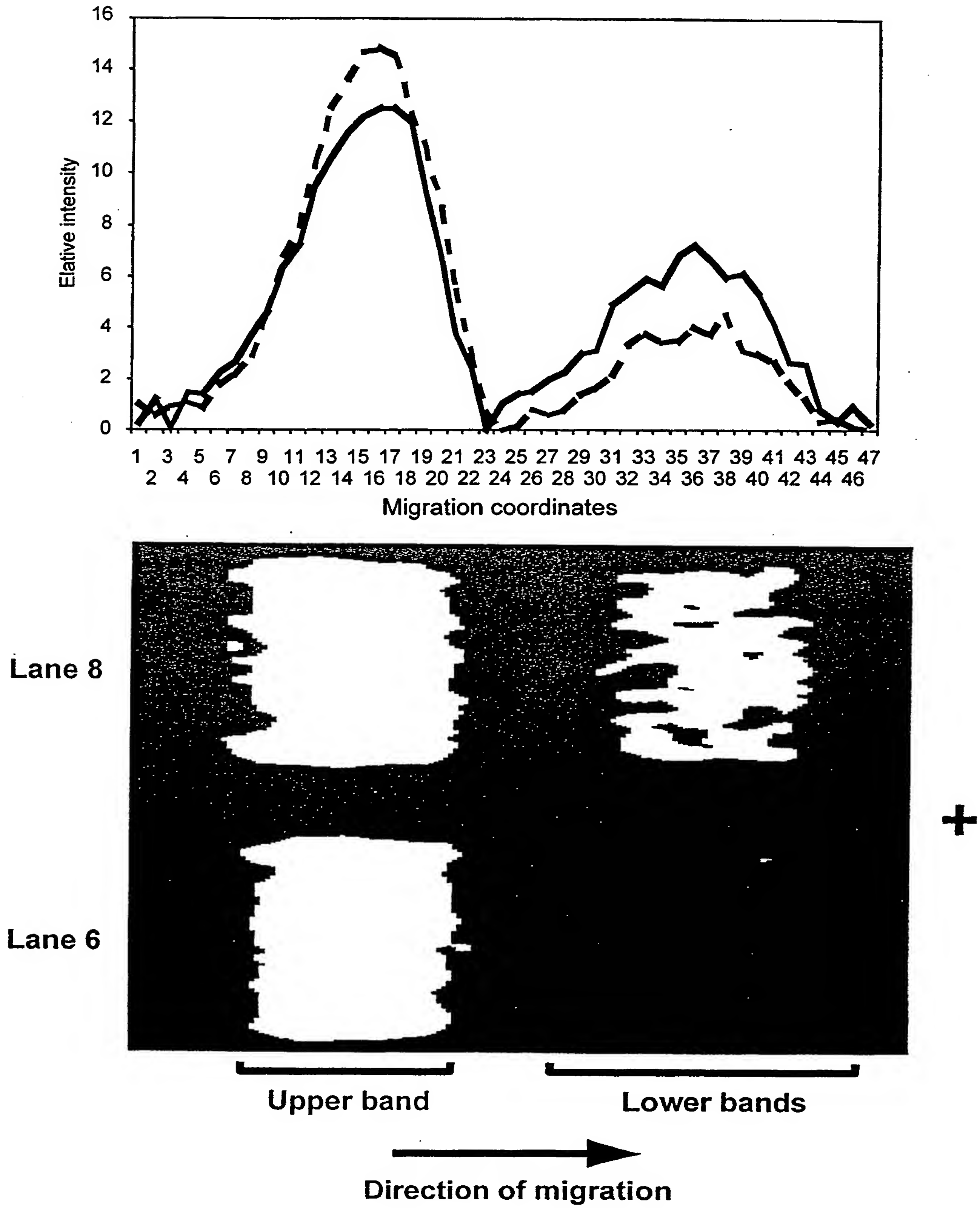

16
14
12
10
8
6
4
2
0
Elative intensity
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46
Migration coordinates
Lane 8
Lane 6
+
Upper band
Lower bands
Direction of migration

FIG. 14

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) **World Intellectual Property Organization**
International Bureau

(43) **International Publication Date**
25 January 2001 (25.01.2001)

PCT

(10) **International Publication Number**
WO 01/05808 A3

(51) **International Patent Classification**[7]: **C12N 15/10**, C12Q 1/68

(21) **International Application Number:** PCT/GB00/02809

(22) **International Filing Date:** 20 July 2000 (20.07.2000)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
9917027.6 20 July 1999 (20.07.1999) GB

(71) **Applicant** *(for all designated States except US)*: **AFFIBODY TECHNOLOGY SWEDEN AB** [SE/SE]; Björnnäsvägen 21, S-113 47 Stockholm (SE).

(71) **Applicant** *(for GB only)*: **GARDNER, Rebecca** [GB/GB]; Frank B. Dehn & Co., 179 Queen Victoria Street, London EC4V 4EL (GB).

(72) **Inventors; and**

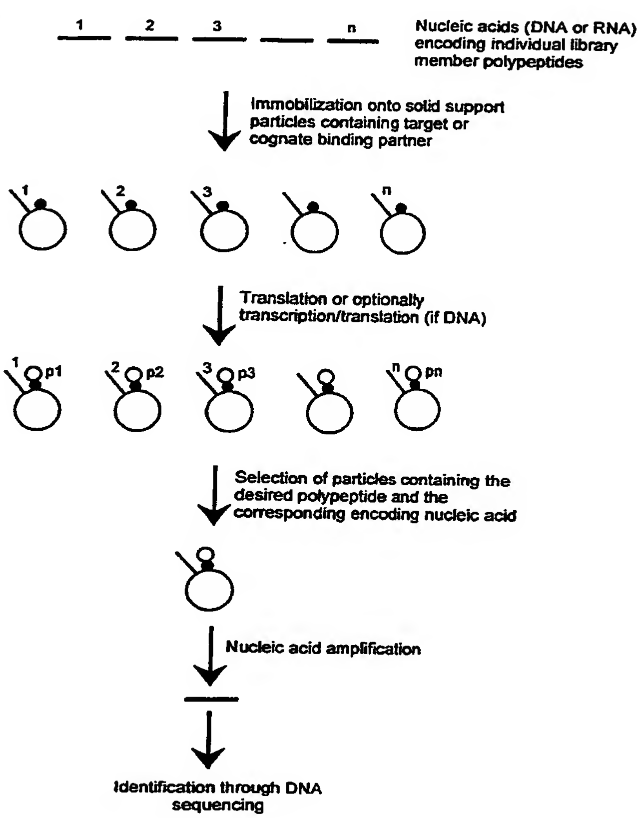
(75) **Inventors/Applicants** *(for US only)*: **NYGREN, Per-Åke** [SE/SE]; Singelgränd 36, S-178 40 Ekerö (SE). **UHLEN, Mathias** [SE/SE]; Måsvägen 8B, S-183 57 Täby (SE). **NORD, Olof** [SE/SE]; Affibody Technology Sweden AB, Björnnäsvägen 21, S-113 47 Stockholm (SE).

(74) **Agents: GARDNER, Rebecca** et al.; Frank B. Dehn & Co., 179 Queen Victoria Street, London EC4V 4EL (GB).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KR (utility model), KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

[Continued on next page]

(54) **Title:** *IN VITRO* SELECTION AND OPTIONAL IDENTIFICATION OF POLYPEPTIDES USING SOLID SUPPORT CARRIERS



(57) **Abstract:** The present invention relates to a method for the selection of one or more desired polypeptides comprising: (a) cell free expression of nucleic acid molecules immobilized on a solid support system to produce polypeptides, the solid support carrying means for biospecific interaction with at least the desired polypeptide or a molecule attached thereto; (b) separation of the solid support carrying both the desired polypeptide and the nucleic acid encoding it; and optionally (c) recovery of the said nucleic acid and/or said desired polypeptide; and molecular libraries for use in such methods.

WO 01/05808 A3

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

(88) Date of publication of the international search report: 2 August 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

Intern[illegible]al Application No
[illegible]/GB 00/02809

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/10 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 99 35293 A (LYNX THERAPEUTICS) 15 July 1999 (1999-07-15) claim 1	1-12, 14-17
A	WO 98 54312 A (BABRAHAM INSTITUTE) 3 December 1998 (1998-12-03) the whole document	1-12, 14-17

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
11 January 2001	24/01/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Masturzo, P

1

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Claims Nos.: 13

Claim 13, referring to every peptide or nucleic acid identified by the technique of claim 1 ff is indefinite as it might cover every known peptide and nucleic acid. This is found in contrast with the requirements of Art. 6 and Rule 6 PCT, and consequently no search for this claim has taken place.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB 00/02809

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9935293 A	15-07-1999	AU 2113999 A	26-07-1999
		EP 1054999 A	29-11-2000
		NO 20003531 A	05-09-2000
WO 9854312 A	03-12-1998	AU 725957 B	26-10-2000
		AU 7666698 A	30-12-1998
		EP 0985032 A	15-03-2000